Analyse van Vragenlijsten met behulp van Item Respons Theorie

S.I.Reniers-Mullié
(9312390)

Begeleiders: dr.ir.B.P. Veldkamp
dr.ir.H.J. Vos

Universiteit Twente
Faculteit Gedragswetenschappen

Analyse van Vragenlijsten met behulp van Item Respons Theorie

S.I.Reniers-Mullié
(9312390)
Hoofddorp, mei 2006

Begeleiders: dr.ir.B.P. Veldkamp
dr.ir.H.J. Vos

Inhoudsopgave

Inhoudsopgave **1**

Inleiding **3**

Motivatie van werknemers **4**

Inleiding *4*

Vaardigheden van de organisatie *4*

Commitment *5*

Persoon-Organisatie fit *6*

Tot Slot *8*

Item Respons Theorie **9**

Inleiding *9*

Dichotome item respons modellen *12*

Het Rasch Model: het 1-parameter logistisch model 12

De logistieke functie: het 2-parameter logistisch model 12

Het 3-parameter logistisch model 13

Polytome item respons modellen *14*

Partial Credit Model (PCM) 15

Generalized Partial Credit Model (G-PCM) 17

Rating Scale Model (RSM) 18

Nominal Response Model (NRM) 19

Het Graded Response Model (GRM) 19

Modified Graded Response Model (M-GRM) 21

Tot slot *22*

Schatten van item- en vaardigheidsparameters **23**

Inleiding *23*

Schatten van itemparameters *23*

Schatten van de vaardigheid *26*

Tot Slot *27*

Methoden **28**

De respondenten *28*

Analyses *29*

Resultaten van het onderzoek **32**

Inleiding *32*

Vaardigheden van de organisatie *32*

Uw Vaardigheden *36*

Uw houding ten opzichte van uw organisatie *41*
Affectief commitment 41
Normatief commitment 45
Continue commitment 49

Uw houding ten aanzien van uw beroep *53*
Affectief commitment 53
Normatief commitment 57
Continue commitment 61

Hoe u vindt dat uw organisatie en u bij elkaar passen *65*

Discussie en Conclusie **72**

Referentielijst **77**

Bijlage variantie analyse **80**

Bijlage output SPSS **84**

Inleiding

De Item Respons Theorie (IRT) is een verzameling modellen die het gedrag van personen die antwoorden geven op vragen (items) uit psychologische en onderwijskundige meetinstrumenten beschrijven en verklaren. Meetinstrumenten zijn bijvoorbeeld testen, vragenlijsten of studietoetsen. Met behulp van deze theorie is de kwaliteit onderzocht van een vragenlijst die tot doel heeft om de sterke en zwakke punten van de organisatie te ontdekken, zodat nieuw beleid daarop afgestemd kan worden. Het betreft een vragenlijst bestaande uit 57 vragen die onderverdeeld zijn in 9 schalen. De vragenlijst is ingevuld door medewerkers uit drie verschillende organisaties, in totaal door 1427 mensen.

De vraag die in deze scriptie beantwoord wordt is of de verschillende schalen van de medewerkerstevredenheid vragenlijst en de individuele vragen bruikbaar zijn.
Om een antwoord te geven op deze vraag is het Graded Respons Model gebruikt om de item karakteristieken zoals item-moeilijkheid en het discriminerend vermogen te schatten. Uit de resultaten daarvan blijkt dat alle 9 geanalyseerde schalen van een voldoende kwaliteit zijn. Het betreft de volgende schalen:
- vaardigheden van de organisatie;
- uw vaardigheden;
- uw houding ten op zichte van de organisatie (affectief, normatief, continue);
- uw houding ten aanzien van uw beroep (affectief, normatief, continue);
- hoe vindt u dat u en de organisatie bij elkaar passen.

Voor wat betreft de individuele vragen blijkt dat in elke schaal 1 of 2 items zitten die weinig informatief zijn en dus kunnen worden weggelaten om tot een nog beter resultaat te komen.

Deze scriptie houdt de volgende indeling aan: Eerst wordt er ingegaan op het begrip motivatie van werknemers: wat vinden werknemers van hun organisatie, wat zijn hun eigen positieve punten, welke vormen van commitment zijn er en tot slot hoe goed of slecht past een werknemer binnen zijn of haar organisatie en beroep? Daarna volgt een overzicht van Item Respons Theorie: wat is het, welke soorten dichotome en polytome modellen zijn er en hoe worden item- en vaardigheidsparameters geschat? In het methoden hoofdstuk wordt vervolgens beschreven wie de respondenten zijn en welke methoden van analyse zijn gebruikt. Deze drie hoofdstukken vormen samen het uitgangspunt voor de analyse in het resultaten hoofdstuk. Daarin zijn de vragen per schaal geanalyseerd en de detailresultaten toegelicht. Afsluitend worden deze resultaten besproken en volgt de conclusie.

Motivatie van werknemers

Inleiding

De vraag waarom werknemers bij een bepaalde organisatie werken en/of blijven, is een veelgestelde vraag. Het succes van een organisatie hangt immers af van de inspanning van de werknemers. De mening van de werknemers kan een raad van bestuur inzicht geven in de sterke en zwakke punten van de organisatie en daarmee helpen bij het maken van beleidsplannen en een goede communicatie voor de toekomst. Een tevredenheidsonderzoek kan je inzicht geven in de redenen, meningen en motivatie van werknemers om te blijven werken voor de organisatie. Een dergelijk onderzoek kan door middel van een vragenlijst uitgevoerd worden, werknemers vullen een vragenlijst in waarop vragen staan als: ‘ik voel me thuis bij deze organisatie’, ze kunnen deze vraag dan beantwoorden met een reeks antwoorden die variëren van ‘niet mee eens’ tot ‘mee eens’.
Werknemers blijven bij een organisatie omdat hun inspanning beloond wordt (financieel of anderzijds). Maar er zijn ook veel onderzoeken die bewijzen dat de motivatie van werknemers nog van andere zaken afhankelijk zijn, zoals emotionele binding en organisatie commitment (van Vuuren et al.,2005).
Hieronder worden drie aspecten beschreven die voor een organisatie van belang zijn. Als eerste de vaardigheden van de organisatie, dan commitment en tot slot persoon-organisatie fit. Het betreft hier de 3 categorieën die in het onderzoek voorkomen.

Vaardigheden van de organisatie

Werknemers kijken vaak naar de kracht van een organisatie als geheel, in aanvulling op hun eigen inbreng in de organisatie. Het externe prestige van een organisatie beïnvloedt de kijk en de gevoelens van werknemers naar de organisatie toe. In de literatuur worden er 3 punten aangehaald als het om collectieve doeltreffendheid gaat:
- niveau van collectiviteit: groepen met een sterk gevoel van gezamenlijke werkzaamheid hebben een positief effect op groepsleden, wat er voor zorgt dat het commitment naar de organisatie toe versterkt wordt (van Vuuren et al.,2005)
- niveau van specificiteit: er wordt een onderscheid gemaakt tussen taakgespecificeerde collectieve werkzaamheid en meer algemene collectieve werkzaamheid. Taakgespecifeerde collectieve werkzaamheid verwijst naar de verwachtingen die mensen hebben over de gezamenlijke kracht om tot een gevraagd resultaat te komen in een specifieke setting, terwijl algemene collectieve werkzaamheid de waargenomen collectieve capaciteit definieert in bredere termen, bijvoorbeeld in termen van hogere strategische organisatiedoelen (Gibson et al.,2000).
- manieren van beoordelen: individueel, gemiddelde, per groep of als organisatie als geheel.

Organisaties die hun werkzaamheid willen verbeteren zullen hun werknemers moeten informeren over de collectieve prestaties. Communicatie is hierin erg belangrijk,

omdat algemene uitvoeringsindicatoren niet voor alle werknemers beschikbaar zijn en het moeilijk zal zijn voor de werknemers om deze indicatoren te evalueren en te begrijpen. Dit kan op verschillende manieren: als eerste transformationeel leiderschap: een charismatisch leider communiceert positieve evaluaties en uit zijn vertrouwen in de gezamenlijke capaciteiten (Shamir, House & Arthur, 1993). Als tweede kan de organisatie, als meer individuele aanpak, mentoren aanstellen binnen de organisatie.

Deze werkzaamheid van de organisatie is in het onderzoek onderzocht door 2 schalen uit de vragenlijst: vaardigheden van de organisatie en vaardigheden van de werknemers zelf.

Commitment

Commitment is een complex concept dat verschillende vormen aan kan nemen. De meest gebruikelijke vorm is het commitment van werknemers aan hun werkgevers, wat meestal aangeduid wordt met organisatie commitment. Maar je hebt ook commitment aan bijvoorbeeld je vakgebied, werk of carrière (Meyer,Allen & Smith,1993).

Mayer en Allen (1991) stelden drie duidelijk te onderscheiden thema's vast binnen de literatuur over organisatie commitment:
- Affectief: commitment als een gevoelsmatige band ten opzichte van de organisatie;
- Continuïteit: commitment als gevolg van het kostenplaatje (voor de werknemer) dat komt kijken als de organisatie verlaten wordt;
- Normatief: commitment als een verplichting naar de organisatie toe om te blijven.

Alle drie de vormen hebben met elkaar gemeen dat commitment de relatie van een werknemer met de organisatie karakteriseert en dat het gevolgen heeft om bij de organisatie te blijven dan wel weg te gaan.
Het gevoel dat werknemers bij elke vorm van commitment hebben is echter wel geheel anders. Werknemers met een sterke affectieve commitment blijven binnen een organisatie omdat ze dat willen, werknemers met een sterke continue commitment blijven binnen een organisatie omdat ze wel moeten/geen andere keus hebben en werknemers met een sterk normatief commitment blijven binnen een organisatie omdat ze zich verplicht voelen te blijven.

Affectieve commitment behelst persoonlijke-, stucturele- en baangerelateerde karakteristieken als ook werkervaringen (Mowday et al.,1982). Mayer en Allen (1991) zeggen dat werkervaringen de meeste invloed hebben op affectief commitment. Werknemers met positieve werkervaringen in de organisatie die overeenkomen met hun eigen verwachtingen en hun basisbehoeften bevredigen, ontwikkelen een sterker gevoel van affectie voor de organisatie dan werknemers met negatieve werkervaringen. Affectief commitment ontstaat dus als een werknemer bevredigend werk heeft en zichzelf kan ontwikkelen. Uit onderzoek blijkt dat deze vorm van commitment het meeste bijdraagt aan werknemers die willen bijdragen aan het succes van de organisatie. Continue commitment ontwikkelt zich als werknemers zich gaan realiseren dat ze veel van zichzelf in de organisatie hebben geïnvesteerd en dat

verloren zou gaan als ze de organisatie verlaten of als ze beseffen dat vergelijkbare banen zeer beperkt zijn. Normatieve commitment ontstaat door de socialisering van een werknemer binnen een organisatie of doordat de werknemer zich verplicht gaat voelen omdat hij/zij van de voordelen van de organisatie geniet (bijvoorbeeld cursussen, bijdrage aan studiekosten), maar ook als een werknemer zich verplicht voelt om in zijn huidige beroep te blijven (omdat het familie traditie is of omdat hij/zij een dure opleiding heeft genoten).

Mayer en Allen vinden dat je het beste inzicht krijgt in de relatie tussen de werknemer en de organisatie als je naar alle drie de vormen van commitment kijkt, ze ontwikkelden daarom een drie-componenten model voor organisatie commitment. Ze veronderstellen dat elke vorm van commitment verandert als gevolg van verschillende ervaringen en dat ze gevolgen hebben voor het gedrag op het werk. Alle drie de vormen van commitment hebben een negatieve invloed op het veranderen van baan. Over het gedrag op het werk zeggen Mayer en Allen (1991) het volgende: affectief en (in mindere mate) normatief commitment zullen een positieve invloed hebben op het werk en continue commitment zal geen of een negatieve invloed hebben. Ze stellen dan ook dat niet alle vormen van commitment gelijk zijn en dat organisaties goed moeten overwegen welke vormen van commitment ze bij willen brengen bij hun werknemers. Een persoon met een hoge affectieve commitment zal waarschijnlijk meer moeite stoppen in de nieuwste ontwikkelingen in het vakgebied (door bijvoorbeeld het bijwonen van conferenties, lezen van vakliteratuur) dan een persoon met een lagere affectieve commitment. Dit geldt ook voor personen met een hoge normatieve commitment, maar kan niet gezegd worden over personen met een hoge continue commitment.

Commitment is het in het onderzoek onderverdeeld in zes schalen: affectief commitment, continue commitment en normatief commitment naar de organisatie toe (samengebracht in de categorie: ‘uw houding ten opzichte van de organisatie’) en affectief commitment, continue commitment en normatief commitment naar het beroep toe (samengebracht in de categorie: ‘uw houding ten aanzien van uw beroep’).

Persoon-Organisatie fit

Sollicitanten kiezen hun baan en/of organisatie op basis van ‘fit’, of ze goed passen binnen de baan/organisatie (Judge & Cable,1997; Saks & Ashforth,1997), recruiters maken hun beslissing door te kijken welke mensen passen in de organisatie (Cable & Judge,1997; Kristof-Brown,2000) en werknemers beslissen of ze bij een organisatie blijven of elders gaan werken op basis van hun idee of ze bij de organisatie passen (Cable & Judge, 1996).
Onderzoek heeft een verschil aangetoond tussen persoon-organsiatie fit en persoon-baan fit.
Bij persoon-organisatie fit gaat het om de overeenstemming tussen de persoonlijke waarden van een werknemer en de cultuur van de organisatie. Bij persoon-baan fit gaat het meer om de overeenstemming tussen de kwaliteiten van de werknemer en de verwachtingen van hem/haar binnen zijn/haar functie (Cable & Judge, 1996;

Kristof-Brown, 2000). Een derde vorm van fit is de behoeftevoorziening van de werknemer. Het gaat hier om de overeenstemming tussen de behoeften van de werknemers en de beloningen die ze van de organisatie krijgen voor de inzet en geleverde prestaties op het werk (Cable, 2002).

Persoon-organisatie fit heeft gevolgen voor de organisatie. Als een werknemer vindt dat hij/zij dezelfde waarden heeft als de organisatie en de andere mensen die bij de organisatie werken, dan voelt hij/zij zich meer betrokken bij de organisatie. Eisenberger et al. (1986) stellen dat organisatie support bestaat omdat werknemers hun organisatie personaliseren. Ook draagt de persoon-organisatie fit mee aan het sociale gedrag onder de werknemers onderling (bijvoorbeeld het helpen van collega's). Persoon-organisatie fit zal er voor zorgen dat werknemers bij de organisatie blijven, immers een werknemer die achter zijn organisatie staat en makkelijk werkt en communiceert met andere werknemers, heeft een sterke band met de organisatie en de collega's wat het moeilijker maakt de organisatie te verlaten (Cable, 2002).

Als er een hoge overeenstemming is tussen de eisen van een baan en de kwaliteiten van de werknemer, dan zal dit leiden tot betere prestaties. Mensen die te hoog boven hun niveau werken zullen vervreemden van hun beroep. Als een werknemer over-gekwalificeerd voor de baan is, zal hij of zij vervreemden van de organisatie. In beide gevallen zal dan naar een andere baan gezocht worden (Cable, 2002).

Bij de derde vorm van fit, de behoeftevoorziening van de werknemer, gaat het om de beloningen van de werkgever voor het goede werk van de werknemer, dit kan worden uitgedrukt in bijvoorbeeld salaris, vrije dagen, interessant en uitdagend werk, promotiekansen, erkenning en positieve werkcondities. Als er een hoge overeenstemming is tussen de behoeften hierin van de werknemer en de organisatie, zal dit bijdragen aan de mate waarin men tevreden is over zijn/haar werk. De behoeftevoorziening van de werknemer hangt ook samen met de tevredenheid over zijn/haar carrière. Mensen investeren tijd en energie in hun carrière om terug te krijgen wat ze willen. Dit kan financieel zijn (hoogte van salaris), sociaal (leuke collega's) of psychologisch (invloed/macht over anderen). Als aan deze verwachtingen is voldaan zullen ze hun carrière eerder als succesvol ervaren dan wanneer dat niet gebeurt.

Alle vormen van 'fit' dragen bij aan de tevredenheid over de baan. Hoe meer overeenstemming er is tussen de waarden (individueel of sociaal) van de werknemers en de waarden van de organisatie, hoe hoger de baantevredenheid is en hoe groter het commitment naar de organisatie toe (Kalliath et al.1999).
Twee theoretische rationalen zorgen voor een positieve relatie tussen waardecongruentie en zowel baantevredenheid als organisatiecommitment. Individuen met gelijke waarden delen dezelfde gedachten die leiden tot dezelfde methoden voor het interpreteren van zaken. Dit zorgt voor vermindering van onzekerheid en verbetering van de persoonlijke relatie. In eenzelfde branche zorgt waardecongruentie tussen werknemers er ook voor dat werknemers weten welke rol

ze hebben en wat ze van elkaar kunnen verwachten. Dit zorgt voor minder conflicten (Kalliath et al.1999).

In het onderzoek staan deze vragen in de categorie: ‘hoe vindt u dat de organisatie en u bij elkaar passen’. Van zowel persoon-organisatie fit, persoon-baan fit en de behoefte voorziening staan 3 vragen in deze categorie. Ze worden als gezamenlijke schaal behandeld.

Tot Slot

In dit hoofdstuk ben ik ingegaan op de motieven van werknemers om bij hun organisatie te blijven werken. Verschillende vormen van commitment, ‘fit’ (passendheid) en vaardigheden van zowel de organisatie als de werknemer zelf zijn besproken. Uit de discussie over de motivatie van werknermers blijkt dat het niet om één begrip gaat, maar dat er meerdere soorten zijn, die ook weer met elkaar samenhangen. De vragen in de vragenlijst hebben betrekking op de genoemde drie onderdelen en zijn hierbij theoretisch toegelicht, zodat de vragen in de vragenlijst beter geïnterpreteerd kunnen worden. Aangezien de vragen betrekking hebben op de verschillende soorten begrippen, ligt het voor de hand om aan te nemen dat de vragen niet op een schaal van betrokkenheid/motivatie gemeten kunnen worden. Maar dat er gedifferentieerd moet worden naar verschillende schalen van medewerker betrokkenheid.

De verschillende soorten van motivatie komen allen terug in de vragenlijst die in het resultaten hoofdstuk geanalyseerd wordt. De analyse is voor een groot deel met behulp van de Item Respons Theorie gedaan, daarom wordt deze theorie in het volgende hoofdstuk besproken.

Item Respons Theorie

Inleiding

Als je iets wilt meten, is het nodig om een meetschaal te hebben. Zo kun je bijvoorbeeld iemands lengte meten met een centimeter, of iemands gewicht met een weegschaal. Maar hoe meet je iets wat niet direct meetbaar is, zoals een bepaalde vaardigheid, attitude of latente vaardigheid. Latente variabelen zijn onobserveerbare entiteiten die observeerbare variabelen, zoals testscores of itemresponsen beïnvloeden. De testscore of het item respons is dan een indicator voor het vaardigheidsniveau van een persoon (Embretson & Reise,2000). Er zijn 2 theorieën die hier iets over zeggen, de klassieke testtheorie en de item respons theorie.

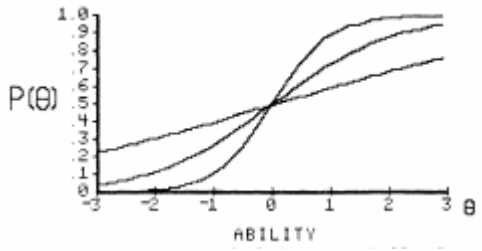
De klassieke testtheorie gaat uit van een toetsscore X die bestaat uit een ware score T en een meetfout E. De testresultaten worden vergeleken met die van een relevante groep personen, namelijk de norm-groep. De basisaannamen van de klassieke test theorie zijn (a) de verwachte waarde van de meetfout is gelijk aan 0 en (b) de meetfout is niet gerelateerd aan andere variabelen (Embretson & Reise,2000). De vaardigheid wordt uitgedrukt door de ware score, die wordt gedefinieerd als: de verwachte waarde van de geobserveerde uitkomst van de test. De vaardigheid van een kandidaat wordt alleen vastgesteld aan de hand van een bepaalde test (Hambleton, Swaminathan & Rogers, 1991). De klassieke test theorie gaat er van uit dat de formule X=T+E altijd van toepassing is. Dit houdt in dat de theorie dus eigenlijk geen theorie genoemd kan worden omdat hij niet gefalsificeerd kan worden. Tekortkomingen van de theorie zijn:
-de klassieke testtheorie kan niet aangeven of het zinvol is een gegeven collectie items tot een toets samen te voegen;
- uit de theorie volgt niet hoe men de items dient te scoren;
- de theorie rechtvaardigt niet dat het zinvol is de itemscores bij elkaar op te tellen (Verhelst,1992).

De Item Respons Theorie (IRT) is een krachtigere theorie dan de klassieke testtheorie en wordt ook wel de moderne testtheorie genoemd (Baker,2001;Verhelst,1992). IRT is een modelgebaseerde meeting waarin vaardigheidsniveau's op basis van een persoon zijn of haar antwoorden en de eigenschappen van de gestelde vragen geschat kunnen worden. IRT wordt steeds meer toegepast vanwege zijn mogelijkheden om praktische meetproblemen op te lossen. Het antwoordpatroon van een persoon op een bepaalde item set voorziet in de basis om het vaardigheidsniveau te schatten (Embretson & Reise,2000).
De IRT werkt met kansen en is gebaseerd op items in plaats van testscore's. De IRT drukt uit hoe groot de kans is dat een item juist beantwoord wordt, als functie van de vaardigheid. Het idee is dat de hoeveelheid vaardigheid die een persoon heeft niet direct waarneembaar is en dat deze alleen verkregen kan worden door de zichtbare antwoorden van personen op een set items. De itemresponsfunctie is dus een conditionele kans en vertelt iets over het gedrag van de persoon als we zijn vaardigheid kennen (Verhelst,1992).

Er zijn veel verschillende vormen van IRT modellen. Zo zijn er IRT-modellen voor dichotome (slechts 2 antwoordcategorieën) en polytome data (meerdere keuzemogelijkheden). Deze scriptie beschrijft de polytome IRT-modellen. De behoefte voor polytome respons formaten is het grootste bij het meten van persoonlijke en sociale variabelen. Omdat de antwoordmogelijkheden meer subtiele nuances bevatten, geven ze meer informatie dan alleen een simpel mee eens/niet mee eens (zoals bij dichotome respons). Dit heeft als voordeel dat de antwoorden meer informatie geven over een groter gebied van het vaardigheidscontinuüm dan dichotome items (Ostini & Nering, 2005).

In de IRT wordt het vaardigheidsniveau van een persoon geschat door de antwoorden op de test items. Een IRT model specificeert hoe zowel vaardigheidsniveau als item kenmerken relateren aan de antwoorden van een persoon. Het vaardigheidsniveau wordt geschat in de context van het model, dus kun je zeggen dat IRT modelgebaseerd meten betreft (Embretson & Reise,2000).
Om de vaardigheid van iemand te meten, heb je een test nodig. Bij een juist antwoord op een item krijg je score 1, en bij een fout antwoord score 0. Elke kandidaat krijgt zo een score die hem ergens op de vaardigheidsschaal plaatst. Deze vaardigheidsscore wordt aangeduid met de letter θ. Op elk vaardigheidsniveau is er een kans dat een kandidaat met een bepaalde vaardigheid een correct antwoord zal geven op de vraag. Deze kans wordt aangeduid met P(θ). De kans op een goed antwoord is bijna 0 voor personen met een zeer lage vaardigheid en voor personen met een hoge vaardigheid loopt deze kans op tot bijna 1. Als je de waarden van P(θ) en de vaardigheid afzet tegen elkaar ontstaat een s-vormige curve, die de item karakteristieke curve wordt genoemd. Elk item in een test heeft zijn eigen unieke curve (Bakker,2001). Een item karakteristieke curve of item karakteristieke functie is een mathematische uitdrukking die de kans op succes (een goed antwoord geven) op een item relateert aan de gemeten vaardigheid door de test en de karakteristieken van het item (Hambleton, Swaminathan & Rogers,1991). Een item karakteristieke curve is onveranderlijk, dit wil zeggen dat de kans op een correct antwoord op een item van een kandidaat niet afhangt van het aantal kandidaten dat hetzelfde vaardigheidsniveau heeft (Hambleton & Swaminathan,1985).

Er zijn -bij het 2-parameter logistisch model- twee eigenschappen van een item karakteristieke curve. De eerste is de moeilijkheid van een item. De tweede de discriminatiegraad, deze beschrijft hoe goed een item onderscheid kan maken tussen kandidaten rond een bepaalde vaardigheid. Hoe steiler de curve in het midden is, hoe beter het item kan discrimineren.
Het punt waar P(θ)=0,5 is de item moeilijkheid (Bakker,2001).



3 item karakteristieke curves met dezelfde moeilijkheid maar met verschillende niveau's van discriminatie

De verzameling van item karakteristieke curves vormen de test karakteristieke curve. Deze curve beschrijft het verwachte aantal goede items op de test op basis van de latente vaardigheid (θ) (Weiss,1995).
Een andere belangrijke functie in de IRT is de informatiefunctie. De functie karakteriseert de precisie van meten voor personen met een verschillende θ. Bij de grafiek wordt de latente vaardigheid θ op de x-as afgezet tegen de hoeveelheid informatie op de y-as. Hoe hoger de informatie des te preciezer is de meeting. De vorm van de item-informatiefunctie hangt af van de itemparameters. Hoe hoger de itemdiscriminatie, hoe gepiekter de informatiecurve zal zijn (Flannery, Reise & Widaman,1995). Verder geldt voor de item-informatiefunctie dat je meer informatie krijgt als de moeilijkheidsparameter dicht bij θ ligt dan wanneer deze waarde verder van θ ligt; je eveneens meer informatie krijgt naarmate de discriminatie parameter hoger is en dat de informatie toeneemt als de c-parameter (de gokkans) naar 0 gaat (Hambleton, Swaminathan & Rogers,1991). Met de aanname van lokale onafhankelijkheid kunnen de item informatie waarden opgeteld worden om zo de test-informatiecurve te vormen (Lord,1980).

Er zijn 2 aanname die gelden bij de IRT:
- de item karakteristieke curves hebben een specifieke vorm;
- er is sprake van unidimensionaliteit en lokale onafhankelijkheid.

De vorm van een item karakteristieke curve beschrijft hoe veranderingen in vaardigheidsniveau relateren aan de kans op een bepaalde respons. Voor dichotome items laat de curve de kans op item-succes zien voor elk vaardigheidsniveau. Bij polytome items toont de curve de kans voor elk vaardigheidsniveau, dat iemand in een bepaalde categorie zal antwoorden.
Unidimensionaliteit wil zeggen dat er maar één vaardigheid gemeten wordt door de items in de test. Hieraan gerelateerd is het concept van lokale onafhankelijkheid. De aanname van unidimensionaliteit kan niet altijd even strikt genomen worden omdat er ook altijd een aantal cognitieve, persoonlijke en testgerelateerde factoren meespelen (zoals motivatie, nervositeit, vermogen om snel te werken, geneigdheid tot gokken). Wat nodig is, is een dominante component of factor die de testresultaten beïnvloedt. Naar deze factor wordt gerefereerd als de vaardigheid die door de test gemeten wordt (Hambleton, Swaminathan & Rogers,1991).
Lokale onafhankelijkheid wordt verkregen als de relatie tussen items of personen volledig gekarakteriseerd wordt door het model. Met andere woorden, lokale onafhankelijkheid wordt verkregen als de kans op het oplossen van een bepaald item onafhankelijk is van de uitkomst van een ander item (Embretson & Reise,2000).

Omdat de polytome modellen uitgaan van de populairste 3 unidimensionele modellen (item respons modellen die één enkele vaardigheid meten): het Rasch model, het 2-parameter logistisch model en het 3-parameter logistisch model, worden deze als eerst beschreven. Dit zijn de dichotome modellen. Daarna volgen de polytome modellen.

Dichotome item respons modellen

Het Rasch Model: het 1-parameter logistisch model

In het Rasch model is de afhankelijke variabele het dichotome antwoord op de vraag (bijvoorbeeld goed – fout) van een bepaalde persoon op een bepaald item. De onafhankelijke variabelen zijn het vaardigheidsniveau (θ) van een persoon en de item moeilijkheid.
Het model wordt ook wel het 1-parameter logistisch model genoemd. Het Rasch model gaat uit van een vaste discriminatie-index van 1. De enige waarde die kan veranderen is de moeilijkheidsparameter. De bijbehorende formule ziet er als volgt uit:

$$P(\theta)= \frac{1}{1+e^{-1(\theta-b)}}$$

Waarbij b de moeilijkheidsparameter (het punt op de moeilijkheidsschaal waar de kans op een goed antwoord 0,5 is) is, en θ het vaardigheidsniveau.
Er zijn verschillende kenmerken van de item karakteristieke curve in het Rasch model. Ten eerste wordt de kans op een goed antwoord groter naarmate de vaardigheid groter is. En ten tweede, items verschillen alleen in moeilijkheid, de helling van de curves zijn gelijk, en kunnen zodoende nooit snijden. Als derde, het buigingspunt van de curve ligt op het punt waar de kans op een goed antwoord op het item 0,5 is.

In het Rasch model is de totale score een voldoende statistiek om het vaardigheidsniveau te schatten, er is geen extra informatie nodig van de data. De specifieke antwoorden die gegeven zijn, voorzien niet in extra informatie over het vaardigheidsniveau. Een antwoordpatroon van
1-1-1-0-0 en een van 1-0-0-1-1, zullen in het Rasch model beide leiden tot score 3, en dus hetzelfde vaardigheidsniveau. Dit komt door de gelijke discriminatie waarde van de items (Embretson & Reise,2000).

De logistieke functie: het 2-parameter logistisch model

De logistieke functie werd voor het eerst afgeleid in 1844 en werd toen gebruikt in de biologie, als model om de groei van planten en dieren weer te geven. Eind jaren 50 werd de functie voor het eerst als model voor de item karakteristieke curve gebruikt (Bakker,2001). De vergelijking voor het 2-parameter logistisch model voegt de item discriminatie parameter toe en wordt dan deze:

$$P(\theta)= \frac{1}{1+e^{-a(\theta-b)}}$$

Net als bij het Rasch model staat de θ voor de vaardigheid en de b voor de moeilijkheid. Het verschil zit hem in de a, die in het 2-parameter logistisch model niet een constante is maar een variabele a.

Het 2-parameter logistisch model is geschikt voor metingen waarbij items gebruikt worden die een verschillend discriminerend vermogen hebben. Anders dan bij het Rasch model kunnen de item karakteristieke curves elkaar dan ook wel snijden en hebben dus niet dezelfde helling (Embretson & Reise,2000).

Het 3-parameter logistisch model

Bij een test met meerkeuzevragen bestaat er altijd de zogenaamde gokkans: de kans dat een kandidaat een item goed maakt, omdat deze het goed gegokt heeft. Het Rasch model en het 2-parameter logistisch model houden hier geen rekening mee. Birnbaum (1968) paste het 2-parameter logistisch model aan en voegde er een derde parameter c aan toe: de gok-parameter (Bakker,2001). De waarde van c varieert niet per persoon, een persoon met een lage vaardigheid heeft dus evenveel kans om het item goed te gokken als een persoon met een hoge vaardigheid. De parameter kan in theorie een waarde aannemen tussen de 0 en 1, maar in praktijk worden waarden hoger dan 0,35 niet acceptabel gevonden. Dit levert de volgende vergelijking op:

$$P(\theta) = c + (1-c)\ \frac{1}{1+e^{-a(\theta-b)}}$$

De item-moeilijkheid is hier nu niet meer (zoals in de vorige 2 modellen) het punt waar de kans op een goed antwoord 0,5 is. Bij het 3-parameter logistisch model is dat punt halverwege de raadkans en 1 ((1+c)/2).

Polytome item respons modellen

De hiervoor beschreven IRT modellen voor dichotome data zijn in veel gevallen beperkt, omdat veel testen of vragenformulieren niet gescoord kunnen worden door een simpel goed of fout. Door voor meer antwoordcategorieën te kiezen, krijg je betrouwbaardere informatie. Polytome IRT modellen zijn nodig om de niet-lineaire relatie tussen de vaardigheid van de kandidaat en de kans op een antwoord in een bepaalde categorie weer te geven (Embretson & Reise,2000). Er zijn geordende polytome items die een expliciete rangorde hebben, zoals bij Likertschalen en er zijn polytome items zonder rangorde, zoals bijvoorbeeld bij meerkeuzevragen (Ostini & Nering,2005). In de volgende tabel staan de modellen kort omschreven met hun belangrijkste kenmerken.

Model	Item Respons Formaat	Model Karakteristieken
Rasch Model / 1 Parameter Logistisch Model	Dichotoom	Discriminatie-parameter is voor alle items gelijk. Moeilijkheids-parameter kan verschillen per item.
2 Parameter Logistisch Model	Dichotoom	Discriminatie- en moeilijkheids-parameters kunnen variëren voor alle items.
3 Parameter Logistisch Model	Dichotoom	Discriminatie en moeilijkheids parameters kunnen variëren voor alle items. Voegt gokkans toe.
Partial Credit Model	Polytoom	Discriminatie-parameter is gelijk voor alle items. Geschikt voor items waar in 'stappen' geantwoord moet worden. Item-moeilijkheidsstappen hoeven niet gelijk te zijn.
Generalized Partial Credit Model	Polytoom	Hetzelfde als Partial Credit Model met als verschil dat de discriminatie parameter kan variëren.
Rating Scale Model	Polytoom	Model is afgeleid van het Partial Credit Model. Discriminatie is gelijk tussen de items. Item-moeilijkheids stappen tussen de items zijn gelijk. Voor elk item moet dezelfde beoordelingsschaal gebruikt worden.
Nominal Model	Polytoom	Respons-categorieën hoeven niet geordend te zijn. Discriminatie-parameter kan variëren tussen de items.
Graded Response Model	Polytoom	Geordende of gerangschikte responses. Discriminatie- parameter kan variëren tussen de items. Geschikt voor testen met verschillende antwoordcategorieën.
Modified Graded Response Model	Polytoom	Geordende of gerangschikte responses. Scheiding tussen de schatting van de item locatie en de categorie tussenruimte parameters. Alle antwoordcategorieën moeten gelijk zijn.

Polytome items zijn categoriale items, ze hebben meer dan twee mogelijke antwoordcategorieën. Categoriale data kunnen beschreven worden in termen van het aantal categorieën waarin de data geplaatst kunnen worden. Geordende categorieën worden gedefinieerd door grenzen of tussenruimtes die de categorieën scheiden. Er is dus altijd één grens of tussenruimte minder dan dat er categorieën zijn.
Vergeleken met dichotome modellen is een belangrijkste probleem met polytome IRT modellen dat het verschil tussen antwoordcategorieën en de grenzen die ze scheiden twee soorten conditionele kansen weergeven. Dit zijn namelijk 1) de kans om in een gegeven categorie te antwoorden en 2) de kans om positief in plaats van negatief te antwoorden op een gegeven grens tussen twee categorieeën. In het dichotome geval komen deze twee kansen op hetzelfde neer. Dat wil zeggen dat de kans om positief dan wel negatief te antwoorden op de categoriegrens (weergegeven door de itemrespons functie) ook de kans om in de positieve categorie te antwoorden weergeeft. Als er meer dan twee categorieën zijn is dit niet langer het geval omdat er altijd minstens 1 categorie is die gedefinieerd wordt door twee grenzen. In dat geval wordt de kans om in die categorie te antwoorden bepaald door de twee categoriegrenzen. Pschychometristen zijn vooral geinteresseerd in de kans om in een gegeven categorie te antwoorden aangezien dat de basis is om een respondent zijn vaardigheidsniveau te bepalen.
Veel polytome IRT modellen werken door naar elke categoriegrens te kijken met een dichotoom model om deze daarna te combineren. Het toepassen van een dichotoom model op elke categoriegrens geeft de kans op een positief antwoord boven een negatief antwoord op die specifieke grens. Het combineren van de dichotome informatie van elke grens geeft de kans op een antwoord in elke individuele categorie. Als je dus wilt weten bij een 5 categorie item wat de kans is om in categorie 2 te antwoorden, moet je kijken naar de grenzen tussen categorie 1 en 2 èn die tussen 2 en 3. De kans om in categorie 2 te antwoorden is dan een combinatie van de kans om positief in de eerste categoriegrens te antwoorden en de kans om negatief in de tweede categoriegrens te antwoorden.
De hierboven beschreven methode wordt gebruikt bij modellen van het Rasch type, zoals het Partial Credit Model en het Rating Scale Model. Ze gebruiken een dichotoom Rasch model item-responsfuctie om de categoriegrenzen te definiëren.
De meer pragmatisch gebaseerde modellen, zoals het Graded Response Model en het Nominal Model gebruiken de 2 parameter logistisch model item respons functie om de categoriegrenzen te definiëren. Alle mogelijke respons categorie antwoorden boven en beneden elke categoriegrens worden bekeken. Deze dichotomisaties kunnen beschreven worden als een set van globale vergelijkingen en behelzen de gehele item categorie respons context op elke grens (Ostini & Nering,2005).

Partial Credit Model (PCM)

Voor items met twee of meer geordende responsen, heeft Masters (1982) binnen het kader van het Rasch model het Partial Credit Model ontworpen. Omdat het model afgeleid is van het Rasch model, beschikt het ook over dezelfde karakteristieken: gescheiden persoons- en item parameter-schattingen voor objectieve vergelijkingen en eenvoudige optellingen zorgen al voor voldoende statistische informatie voor het meten van het vaardigheidsniveau (Masters & Wright,1997). Het PCM is een geordend polytoom Rasch model wat gebaseerd is op de dichotomisatie van aangrenzende categorieën. Het model vereist echter geen geordende categoriegrenzen. Dit resulteert in een separate

locatieparameter (δ) voor elke categoriegrens (j) van elke item (i). Deze benadering staat toe dat het aantal categorieën varieert van item tot item in een test, zelfs voor items met hetzelfde aantal categorieën mogen de grenslocaties variëren (Masters,1982).
Het PCM is dan ook uitermate geschikt voor items waar antwoorden deels goed kunnen zijn, zoals bijvoorbeeld in wiskundige testen. Het PCM is ook erg geschikt voor attitude of persoonlijkheids testen, waar door middel van een meerpuntsschaal geantwoord moet worden op stellingen.
Het PCM kan worden gezien als een uitbreiding op het 1-parameter logistisch model en wordt als volgt geschreven:

$$P_{ix}(\theta) = \frac{\exp\left[\sum_{j=0}^{x} (\theta - \delta_{ij})\right]}{\sum_{r=o}^{mi}\left[\exp \sum_{j=0}^{r} (\theta - \delta_{ij})\right]}$$

δ_{ij} wordt ook wel de moeilijkheidsstap genoemd, hoe hoger de waarde van een bepaalde δ_{ij} hoe moeilijker een bepaalde stap is ten opzichte van een andere stap binnen dat item. Het geeft de relatieve moeilijkheid van elke stap aan. De δ_{ij} term kan ook gezien worden als het punt op de latente vaardigheidsschaal waar de respons curves van 2 opvolgende categorieën elkaar kruisen.
Van alle items wordt verondersteld dat ze dezelfde steilheid van curve hebben, net als in het Rasch model wordt de discriminatie index dus gelijk geacht.
Een voorbeeld van een 4 categorie attitude item:

0----------1----------2----------3
helemaal niet mee eens een beetje mee oneens een beetje mee eens mee eens
stap 1 *stap 2* *stap 3*

In dit item moeten kandidaten 3 stappen doorlopen om in de hoogste categorie te kunnen antwoorden, namelijk besluiten tussen helemaal niet mee eens en een beetje mee oneens (stap 1), daarna tussen een beetje mee oneens en een beetje mee eens (stap 2), en tot slot tussen een beetje mee eens en mee eens (stap 3). Sommige stappen zullen makkelijker te maken zijn dan anderen. De δ_{ij} parameters geven aan waar op het latente vaardigheid continuüm de categorie-responscurve kruist en geeft zo dus aan waar op de latente vaardigheidsschaal het antwoord van een bepaalde categorie meer waarschijnlijk wordt dan dat van een vorige categorie (Embretson & Reise,2000).

Masters en Wright (1997) stellen dat de eenvoudigheid van het PCM zijn grootste kracht is, wat komt door het toepassen van een dichotoom Rasch model voor elk paar van aangrenzende item categorieën. Het resultaat is een polytoom IRT model met slechts

twee lokatieparameters: een voor items δ_{ij} en een voor personen $\theta_{j.}$ Deze parameters zijn gescheiden, wat er voor zorgt dat er zeer objectief vergeleken kan worden binnen en tussen persoonsparameters en itemparameters.

Een andere manier om de relatie tussen de vaardigheid van de kandidaat en het item respons weer te geven is om het PCM te gebruiken om de verwachte of ware score op een item weer te geven in een curve. Deze curve geeft dan de verwachte score van een kandidaat weer met een bepaalde vaardigheid. Als deze curves opgeteld worden (per item) dan geeft de opgetelde curve de verwachte score aan voor alle kandidaten met een bepaalde vaardigheid (Embretson & Reise,2000).
De curve kan met de volgende formule berekend worden:

$$E(X) = \sum_{x=0}^{mi} x\, P_x(\theta)$$

Waarbij x = categorie 0,1,……,m voor elk item
$P_x(\theta)$= de kans om in categorie x te antwoorden met een gegeven θ

Generalized Partial Credit Model (G-PCM)
Het Generalized Partial Credit Model (Muraki,1992) is een generalisatie van het PCM, die toestaat dat de discriminatieparameter varieert tussen de items. Muraki stelt dat door het toestaan van variatie in de discriminatie tussen alle items in een test er meer inzicht is in de karakteristieken van de test-items. De formule is gelijk aan die van het PCM, maar er is een discriminatieparameter α aan toegevoegd:

$$P_{ix}(\theta) = \frac{\exp\left[\sum_{j=0}^{x} \alpha_i(\theta - \delta_{ij})\right]}{\sum_{r=o}^{mi}\left[\exp \sum_{j=0}^{r} \alpha_i(\theta - \delta_{ij})\right]}$$

De moeilijkheids-stap-parameter δ_{ij} wordt in dit model hetzelfde geïnterpreteerd als in het PCM. Namelijk het snijpunt van 2 curves van aangrenzende categorieën. Dat zijn de punten op de latente vaardigheidschaal waar de ene respons categorie meer aannemelijk wordt dan de andere. De discriminatieparameter α wordt echter niet op dezelfde manier geïnterpreteerd. In polytome modellen hangt de itemdiscriminatie af van de discriminatie-parameter en de spreiding van de categoriedrempels (in GRM of M-GRM) of categorie-tussenruimtes (in PCM of G-PCM). In het G-PCM geven de discriminatieparameters de mate aan waarin categoriale responsen veranderen tussen items als het vaardigheidsniveau verandert (Embretson & Reise,2000).

Rating Scale Model (RSM)

Het Rating Scale Model (Andrich,1978) is ook een lid van de Rasch familie. Het model is afgeleid van het Partial Credit Model, met dezelfde beperking van gelijke discriminatie voor alle items. Items die gebruikt worden in psychologische context zoals attitude, interesse of persoonlijkheidsonderzoeken hebben vaak dezelfde structuur. Meestal is dit een Likertschaal achtig type, waar mensen gevraagd wordt op een item te antwoorden middels een voorgedefinieerde set van responsen en waar dezelfde set van responsalternatieven toegepast wordt in de gehele test. Dus, in tegenstelling tot bij het PCM, moeten alle items in een gegeven setting hetzelfde aantal categorieën hebben en de categorieën moeten dezelfde set labels hebben (Ostini & Nering,2005).

Het verschil in beide modellen zit in de afstand tussen de moeilijkheids stappen (of niveau's) van categorie tot categorie binnen elk item, want deze is hetzelfde voor alle items in het RSM, in tegenstelling tot het PCM waar geen aannamen gemaakt zijn over de relatieve moeilijkheden van de stappen binnen een item. Bij een wiskundige som bijvoorbeeld, is het heel logisch dat sommige stappen om bij het uiteindelijke antwoord te komen, makkelijker zijn dan andere. Als er echter met een attitude-vragenlijst (niet mee eens-neutraal-mee eens) gewerkt wordt, is het niet erg aannemelijk dat de verschillende stappen die gemaakt moeten worden binnen het item verschillen in moeilijkheid.

In het RSM vallen de stapmoeilijkheden uiteen in twee componenten, namelijk λ_i en δ_j waarbij δ_{ij} gelijk is aan $\lambda_i + \delta_j$. De λ_i parameter is een item-locatieparameter die voor elk individueel item in de test geschat wordt. δ_j zijn de categoriegrenzen, die in het RSM aangeduid worden als tussenruimte parameters. Deze geven de grenzen tussen de categorieën van de rangschaal weer, gerelateerd aan de vaardigheidslocatie van elk item. De δ_j vertelt je dus hoever elke categoriegrens afligt van de locatieparameter. Deze parameters worden in een keer geschat voor de gehele set van items, dit houdt wel in dat de tussenruimteparameters geordend moeten zijn. Het RSM heeft dus minder parameters die geschat moeten worden (n item locatie parameters + m tussenruimte- parameters (categoriegrenzen) versus n x m item (categoriegrenzen) parameters) dan voor het PCM (Ostini & Nering,2005).

De responsfunctie kan als volgt geschreven worden:

$$Px(\theta) = \frac{\exp\left\{\sum_{j=0}^{x}[\theta - (\lambda_i + \delta_j)]\right\}}{\sum_{x=0}^{M}\exp\left\{\sum_{j=o}^{x}[\theta - (\lambda_i + \delta_j)]\right\}}$$

Het RSM gaat uit van een vaste set van schaalpunten die gebruikt worden voor de gehele item set. Als de items binnen de schaal verschillende formaten hebben, dan is het RSM dus geen geschikte keuze (Embretson & Reise,2000).

Nominal Response Model (NRM)

Bock (1972) heeft een model gemaakt waarbij responsen niet langs het vaardigheidscontinuüm geordend hoeven te zijn. Bij het NRM zijn de itemresponsen gecategoriseerd, maar de categorieën zijn niet geordend, hierdoor kunnen de scores niet bij elkaar opgeteld worden om tot een totale score te komen (Baker & Kim,2004). Het model is oorspronkelijk ontworpen om de afleidende antwoorden bij meerkeuze vragen te karakteriseren, maar het model kan gebruikt worden voor alle items waar geen voorgespecificeerde antwoordcategorieën voor bepaald zijn (Embretson & Reise,2000). Het doel van het model is om de precisie van de verkregen geschatte vaardigheid te maximaliseren door de verkregen informatie uit elk item te gebruiken. Elke itemoptie wordt beschreven in een item-optie-karakteristieke-curve, zelfs de weggelaten antwoorden kunnen in een curve weergegeven worden. De curve voor het juiste antwoord zal een monotoon stijgende zijn. Voor de onjuiste opties zal de vorm van de curve afhangen van de waargenomen antwoorden van de kandidaten met verschillende vaardigheidniveau's (Hambleton & Swaminathan,1985).

In het NRM kan de kans dat een respondent in categorie x antwoord als volgt geschreven worden:

$$P_{ix}(\theta) = \frac{\exp(\alpha_{ix}\,\theta + c_{ix})}{\sum_{x=o}^{m} \exp(\alpha_{ix}\,\theta + c_{ix})}$$

In dit model moeten een α_{ix} en een c_{ix} parameter geschat worden voor elke van de $m_i + 1$ antwoordcategorieën binnen een item. De α_{ix} staat voor de discriminatieparameter en de c_{ix} is een interceptparameter voor categorie x.

Het NRM is het enige model dat voorziet in een middel voor de gelijktijdige schatting van de item parameters van alle responsalternatieven van een meerkeuze item. Het model kan ook de vaardigheid van een kandidaat schatten als alle respons categorieën in acht genomen worden, zelfs als een onuitgewerkte testscore niet mogelijk is (Baker & Kim,2004).

Het Graded Response Model (GRM)

Voor vragen met 3 of meer antwoordcategorieën heeft Samejima (1969) een model gemaakt voor gerangschikte of geordende responsen. Een antwoord kan gerangschikt zijn in een reeks van scores, bijvoorbeeld van zeer slecht (0) tot uitstekend (9). Of in een onderzoek kan een proefpersoon kiezen uit een aantal geordende opties, als bijvoorbeeld de Likertschaal (zie voorbeeld hieronder) die loopt van *helemaal mee oneens* tot *helemaal mee eens* (Mellenberg,1994).

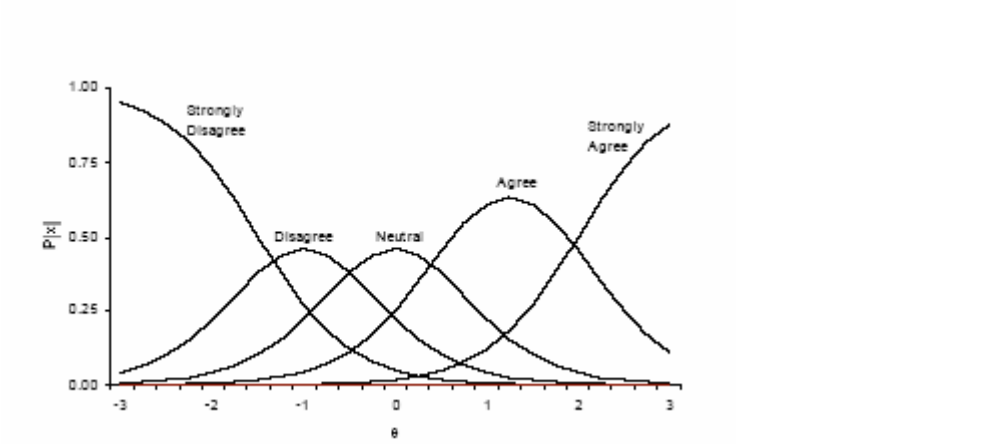
Score x	helemaal mee oneens	mee oneens	neutraal	beetje mee eens	mee eens
	0	1	2	3	4

Tussenruimte j	1	2	3	4

Het GRM is een uitbreiding van het 2-parameter logistisch model en geeft de kans dat een respondent zal antwoorden in een categorie k of hoger. Het GRM werkt ook met 2 itemparameters om itemeigenschappen weer te geven. Bij de GRM worden de items gedefinieerd door een parameter die de helling aangeeft -de discriminatie parameter (α, deze kan variëren per item, maar is binnen een item constant)- en door 2 of meer locatieparameters (β). (Het aantal locatieparameters is gelijk aan het aantal antwoordcategorieën min één.)
De locatieparameters geven de tussenruimtes aan van de geordende antwoordcategorieën langs de θ-schaal weer. De β-parameters representeren het vaardigheidsniveau dat nodig is om boven die categorie te antwoorden met 50% kans. Dus kandidaten met een vaardigheid hoger dan de locatieparameter hebben 50% kans om in die of een hogere categorie te vallen.
De eerste stap in het schatten van de waarschijnlijkheden in de GRM is het berekenen van de curves die de waarschijnlijkheid weergeven dat het gegeven antwoord in of boven de gegeven categorietussenruimte valt onder voorwaarde van het vaardigheidsniveau θ. Dit gebeurt door middel van de 2-parameter logistische functie:

$$P^*_{ix}(\theta) = \frac{\exp\left[\alpha_i\,(\theta - \beta_{ij})\right]}{1 + \exp\left[\alpha_i\,(\theta - \beta_{ij})\right]}$$

Als deze geschat zijn, kunnen de werkelijke waarden berekend worden. Dit kan als volgt:
$P_{ix}(\theta)=P^*_{ix}(\theta)-P^*_{i(x+1)}(\theta)$
De kans dat in of boven de laatste categorie geantwoord wordt is 1 en de kans om in de hoogste antwoordcategorie te antwoorden is gelijk aan 0.
De kans van een kandidaat, gegeven een vaardigheid, om in een bepaalde categorie te antwoorden kan dan weergegeven worden in een categorie respons curve, ook wel item karakteristieke curve genoemd. Deze curves geven de waarschijnlijkheid weer om in een bepaalde antwoordcategorie te antwoorden gegeven het vaardigheidsniveau. In het algemeen geldt bij deze curves dat hoe hoger de hellingparameters (discriminatie parameters: α) des te smaller en gepiekter de curves zijn, wat aangeeft dat de antwoordcategorieën goed onderscheid maken tussen de vaardigheidsniveau's. De locatieparameters (β) zeggen iets over de breedte van de curves en hoe ver de categorie tussenruimten van elkaar afliggen (Embretson & Reise,2000).
De figuur hieronder laat een voorbeeld zien van een geordend item met 5 antwoord-categorieën (stongly diasagree-disagree-neutral-agree-strongly agree). Het model geeft aan hoe aannemelijk het is dat een persoon met een bepaalde vaardigheid θ een bepaalde antwoordcategorie zal kiezen.



Item karakteristieke curves voor een item met 5 gegradeerde antwoordcategorieën.

Wat opvalt aan de item karakteristieke curves is dat ze niet dezelfde vorm hebben. De uiterste antwoorden (stongly disagree en strongly agree) hebben een respectievelijk monotoon dalende en monotoon stijgende functie. De drie niet-extreme categorieën hebben een item karakteristieke curve die niet monotoon is, dat wil zeggen dat de kans om in die categorie te antwoorden toeneemt tot een bepaald punt naarmate de vaardigheid (θ) toeneemt en dan –vanaf dat punt- weer afneemt.
Het GRM zorgt voor hogere item- en test-informatiefuncties dan bij dichotoom gescoorde items zou worden verkregen. De vaardigheid kan dus met een grotere precisie geschat worden of hetzelfde niveau van precisie kan bereikt worden door minder items te gebruiken (Baker & Kim,2004).

Modified Graded Response Model (M-GRM)
Muraki (1990) heeft een wijzigbaar GRM ontworpen dat het mogelijk maakt het model te gebruiken bij vragenlijsten waarbij in classificaties geantwoord moet worden. Net als bij het GRM, staat het M-GRM toe om de discriminatie van de itemparameters te variëren tussen de items. Maar in de M-GMR wordt de β-parameter onderverdeeld in 2 termen, namelijk: een locatieparameter (b_i) voor elk item en een set van categorie-tussenruimte-parameters (c_j) voor de gehele schaal. Dan is $\beta_{ij}=b_i-c_j$. De curves voor het M-GRM kunnen dan als volgt geschreven worden:

$$P^*_{ix}(\theta)= \frac{\exp \alpha_i(\theta-b_i+c_j)}{1+ \exp \alpha_i(\theta-b_i+c_j)}$$

De kans om te antwoorden in een bepaalde categorie is dan :
$P_{ix}(\theta)=P^*_{ix}-P^*_{i(x+1)}$

De M-GRM is beperkt, want het model neemt aan dat de categoriegrenzen gelijke afstanden van elkaar hebben terwijl dat bij de GRM niet hoeft. Een voordeel van de M-

GRM boven de GRM is echter dat het model een scheiding van de schatting van een itemlocatie en categorietussenruimtes toestaat. Een voordeel hiervan is dat de item-locatieparameters (b_i) gebruikt kunnen worden om de items in volgorde van moeilijkheid (of schaalwaarde) te zetten (Embretson & Reise,2000).

Tot slot

In dit hoofdstuk ben ik ingegaan op de Item Respons Theorie. De theorie is vergeleken met de klassieke testtheorie en de verschillende soorten dichotome en polytome modellen zijn besproken. Omdat de vragenlijst uit meerdere (7) geordende antwoordcategorieën bestaat wordt het Graded Respons Model van Samejima (1969) gebruikt voor de analyse. Het GRM is een uitbreiding van het 2-parameter logistisch model en geeft de kans weer dat een respondent zal antwoorden in een categorie k of hoger.

Voordat de respondenten, de methoden en de resultaten van het onderzoek besproken zullen worden, volgt eerst nog een hoofdstuk over het schatten van de item- en vaardigheidsparameters.

Schatten van item- en vaardigheidsparameters

Inleiding

Omdat het schatten van item-en vaardigheidsparameters voor polytome IRT modellen buiten het bereik van deze scriptie valt wordt in dit hoofdstuk alleen het schatten van item- en vaardigheidsparameters behandeld voor de dichotome item respons modellen.

Om de vaardigheid van een persoon te schatten wordt er gekeken naar de antwoorden op de test. Maar om het vaardigheidsniveau te vinden aan de hand van het antwoordenpatroon is meer een zoekproces dan een scoreprocedure. Dat houdt in dat er wordt gezocht naar het vaardigheidsniveau dat het meest waarschijnlijk is gegeven de gegeven antwoorden. Vaardigheidsniveau's worden geschat door de grootste-aannemelijkheidsmethode: het geschatte vaardigheidsniveau van een persoon maximaliseert de waarschijnlijkheid van zijn of haar antwoordpatroon gegeven de item responsen. Dus om een geschikt vaardigheidsniveau te vinden moeten de waarschijnlijkheden van antwoordpatronen onder verschillende vaardigheidsniveau's worden weergegeven en het vaardigheidsniveau met de grootste waarschijnlijkheid moet gevonden worden.
Om de meest waarschijnlijke vaardigheidsscore te vinden moet eerst de waarschijnlijkheid van een persoons antwoordenpatroon in een model dat de eigenschappen van de items weergeeft, uitgedrukt worden. Als dat gedaan is kan de waarschijnlijkheid van een persoons antwoordenpatroon voor elk hypothetisch vaardigheidsniveau uitgerekend worden. Daarna kunnen de waarschijnlijkheden afgezet worden tegen de vaardigheidsniveau's (Embretson & Reise,2000).
Alleen de vaardigheidsparameter θ wordt gekarakteriseerd door de kandidaat, de overige parameters (α, β en c) karakteriseren de items. De item- en vaardigheidsparameter zijn meestal onbekend. Om deze te achterhalen wordt er gekeken naar de antwoorden op de items van de kandidaten. Gegeven deze antwoorden worden de parameters geschat (Hambleton & Swaminathan,1985).
In de volgende paragrafen wordt ingegaan op het schatten van de item- en vaardigheids-parameters voor dichotome item respons modellen.

Schatten van itemparameters

Item parameters worden meestal geschat door de grootste-aannemelijkheidsmethode. De meest gebruikte methodes zijn de 'joint-maximum-likelihood' (JML), de 'marginal-maximum-likelihood' (MML) en de 'conditional-maximum-likelihood'(CML). Ze verschillen in de manier waarop de waarschijnlijkheden van de geobserveerde antwoordpatronen worden geconceptualiseerd. Deze methodes zijn gebaseerd op de maximum-likelihood principes. Deze principes specificeren hoe fouten in de modelvoorspelling gekwantificeerd worden. Fouten (niet-waarschijnlijke geoberserveerde data) moeten gekwantificeerd worden omdat de parameters in het model geschat worden door de fout zo klein mogelijk te laten zijn. Bij grote steekproeven hebben maximum-likelihood schattingen van parameters verschillende voordelen: ze zijn consistent (convergentie van de waarde als

de steekproefgrootte toeneemt), efficiënt (relatief kleine standaardfout) en de schattingsfout een standaard normale verdeling heeft (Embretson & Reise,2000).
De parameters die het gekozen item respons model karakteriseren moeten geschat worden omdat ze onbekend zijn. In item respons theorie hangt de kans van een goed antwoord af van de vaardigheid van een kandidaat en van de discriminatie- en moeilijkheidparameters. Het enige wat bekend is, zijn de antwoorden op de test. De parameters worden geschat middels de grootste-aannemelijkheidsmethode. De waarschijnlijkheidsfunctie van N kandidaten die n items beantwoorden is (met de aanname van lokale onafhankelijkheid):

$$L(u_1,u_2,\ldots\ldots\ldots u_N \mid \theta\ ,\ a,b,c) = \prod_{i=1}^{N} \prod_{j=1}^{n} P_{ij}^{u_{ij}}\ Q_{ij}^{1-u_{ij}}$$

Waarin:
u_i : antwoordenpatroon van kandidaat i op n items;
θ : vector van N vaardigheidsparameters;
a,b,c: vectors van item parameters voor de n-item test;
P_{ij} = P(antwoord j uit vector $U_i|\theta$, a,b,c) en Q_{ij} = 1-P(antwoord j uit vector $U_i|\theta$,a,b,c).

De waarde van de item- en vaardigheidsparameters die de waarschijnlijkheidsfunctie maximaliseren kunnen zo geschat worden. In de ‘joint-maximum-likelihood-estimation’ procedure (joint geeft aan dat er twee soorten parameters tegelijkertijd geschat moeten worden, persoonsparameters en itemparameters) wordt deze beslissing in twee fases genomen. In de eerste fase worden aanvangswaarden voor de vaardigheidsparameter gekozen. De logaritme van het aantal goed gescoorde items op het aantal fout gescoorde items voor elke kandidaat voorziet in een goede startwaarde. Deze vaardigheidswaarden worden dan als bekend beschouwd en dan worden de itemparameters geschat.
In de tweede fase worden de itemparameters als bekend beschouwd en worden de vaardigheidsparameters geschat. Deze procedure wordt herhaald tot de waarden van de schattingen niet meer veranderen tussen de twee opvolgende schattingfases (Hambleton, Swaminathan & Rogers,1991).
De JML procedure heeft enkele voordelen, namelijk dat het algoritme makkelijk te programmeren is en dat het van toepassing is op veel IRT-modellen (Embretson & Reise,2000).
Deze ‘joint-maximum-likelihood’ heeft echter ook een aantal nadelen:
- vaardigheidschattingen met perfecte en 0-scores bestaan niet;
- itemparameterschattingen voor items die allemaal correct (of incorrect) gescoord zijn door de kandidaten bestaan niet;
- in het 2- en 3-parameter model levert de procedure geen consistente schattingen op voor de item- en vaardigheidsparameters;
- in het 3-parameter model kan de procedure voor het vinden van de schattingen falen.
Het probleem van de consistentie houdt in dat hoe meer informatie men verzamelt over een parameter door een steekproef groter te maken, des te nauwkeuriger de schatting moet zijn, en in de limiet, bij n die nadert naar oneindig is de kans dat men de parameter juist schat gelijk aan 1. Er treedt dan een complicatie op: om meer informatie te verzamelen over itemparameters dient men de toets steeds bij nieuwe personen af te

nemen, maar elke persoon die men aan de steekproef toevoegt brengt zijn eigen onbekende θ-parameter mee. Dit wil zeggen dat de omvang van het probleem, het aantal te schatten parameters, even snel groeit als het aantal personen in de steekproef.

De problemen met de ‘joint-maximum-likelihood’ procedure ontstaan omdat de item- en vaardigheidsparameter gelijktijdig geschat worden. Dit probleem zou verdwijnen als de itemparameters geschat zouden worden zonder verwijzing naar de vaardigheidsparameters. Als de kandidaten beschouwd worden als een willekeurig gekozen groep uit de populatie, dan kunnen ze - door een verdeling van de vaardigheidsparameters te specificeren - uit de waarschijnlijkheidsfunctie geïntegreerd worden. Dit resulteert in een ‘marginal-maximum-likelihood’ schatting (Hambleton, Swaminathan & Rogers,1991). In de MML schatting worden de onbekende vaardigheidsniveau’s behandeld door de antwoordpatroonmogelijkheden uit te drukken als verwachtingen van een populatieverdeling. Men gaat er vanuit dat het niet uitmaakt wie er in de steekproef zit, omdat je alleen geïnteresseerd bent in de itemparameters. Dit impliceert dat de steekproef als een aselecte steekproef uit een of andere populatie beschouwd kan worden en dat de gedragingen van die toevallige steekproef gebruikt kan worden om de itemparameters te schatten (Embretson & Reise,2000). Voor een goede schatting van de vaardigheidsparameter is een grote groep van kandidaten belangrijk, dus deze procedure kan alleen gebruikt worden als je een voldoende grote groep kandidaten tot je beschikking hebt. De itemparameters kunnen nu geschat worden en dan als bekend beschouwd worden (Hambleton, Swaminathan & Rogers,1991).
De MML vereist dat de verdeling van vaardigheidsniveau’s gespecificeerd is. De verdeling wordt gespecificeerd als normaal, maar de verdeling hoeft niet op voorhand al bekend te zijn. Als de steekproef groot genoeg is, kan de vaardigheidsverdeling geschat worden uit de data.
De MML schattingsmethode heeft veel voordelen:
- deze is toepasbaar op alle IRT modellen;
- deze is bruikbaar voor zowel lange als korte testen;
- de schattingen van item standaard fouten worden beschouwd als goede benaderingen van de verwachte steekproefvariantie van de schattingen;
- ook perfecte scores (alles goed) kunnen geschat worden.
Er zijn slechts twee kleine nadelen te noemen, namelijk dat het effectieve algoritme moeilijk te programmeren is en dat er een verdeling voor het vaardigheidsniveau moet worden aangenomen (Embretson & Reise,2000).

Een ander methode om de θ-parameters kwijt te raken is de ‘conditional-maximum-likelihood’ (er wordt geconditioneerd op de score) schatting. CML schattingen kunnen alleen voor het Rasch-model (en aanverwante modellen) gebruikt worden omdat CML vereist dat de totale score voldoende moet zijn voor het schatten van het vaardigheidsniveau. De totale steekproef wordt hierbij opgedeeld in homogene scoregroepen, dat wil zeggen in groepen van personen die een zelfde aantal items correct hebben (ongeacht welke items goed zijn gescoord), waarna de aannemelijkheid van een bepaald antwoordpatroon bekeken wordt binnen elke scoregroep. Er wordt bij de ‘conditional-maximum-likelihood’ procedure geen enkele veronderstelling gemaakt over de verdeling van θ, in tegenstelling tot bij de ‘marginal-maximum-likelihood’ procedure (Hambleton, Swaminathan & Rogers,1991).

Ook aan dit model kleven weer verschillende voor- en nadelen. Voordelen van CML schattingen zijn:
- een voorafgestelde verdeling van het vaardigheidsniveau is niet nodig;
- de itemparameterschattingen reflecteren het principe van invariante item parameters: de schattingen zijn invariant omdat ze niet direct beïnvloed worden door het vaardigheidsniveau van de kandidaten in de steekproef (dit wordt ook wel het principe van specifieke objectiviteit genoemd);
- de schattingen zijn consistent en efficiënt en normaal verdeeld.

De nadelen zijn dat de schattingen alleen bij modellen van het Rasch type gemaakt kunnen worden, er geen schattingen gemaakt kunnen worden voor kandidaten met een perfecte score en dat er numerieke problemen ontstaan bij langere testen (Embretson & Reise, 2000).

Schatten van de vaardigheid

IRT modellen worden gebruikt om een persoon zijn of haar vaardigheid te schatten door eerst de waarschijnlijkheid van een antwoordenpatroon te schatten, gegeven het niveau van de onderliggende vaardigheid die door de schaal gemeten wordt. Omdat de items lokaal onafhankelijk zijn, is de waarschijnlijkheidsfunctie L:

$$L = \prod_{i=1}^{n\ items} T_i\,(u_i \mid \theta)$$

Deze is het product van de individuele item respons functies: $T_i\,(u_i \mid \theta)$ staat voor de kans op een antwoord u op het item i gegeven de onderliggende vaardigheid θ.
Daarna worden vaardigheidsniveau's geschat door de grootste-aannemelijkheidmethode; de vaardigheid van een kandidaat maximaliseert de waarschijnlijkheidsfunctie gegeven de item eigenschappen. Een kandidaat zijn of haar vaardigheidsniveau wordt dus geschat door een proces dat 1) de waarschijnlijkheid van een antwoordpatroon berekent voor de niveau's van de onderliggende vaardigheid en 2) een zoek methode gebruikt om het vaardigheidsniveau te vinden dat het meest waarschijnlijk is (Embretson & Reise,2000).

De maximale waarschijnlijkheidsprocedure begint met vooraf gestelde waarden voor de vaardigheid van de kandidaat en de bekende waarden van de itemparameters. Deze worden gebruikt om de kans op een correct antwoord op elk item van een kandidaat te berekenen. Dan wordt een aanpassing gemaakt op de vaardigheidsschatting die de overeenkomst van de berekende kansen met de kandidaat zijn itemresponsvector verbetert. Het proces wordt net zo lang herhaald totdat de aanpassingen klein genoeg zijn om aan te nemen dat het resultaat de schatting is van de kandidaat zijn vaardigheidsparameter (Baker,2001).

Het responspatroon van een random gekozen kandidaat op een set van n items ziet er als volgt uit: $u_1,u_2,\ldots\ldots,u_n$, waar de u kan staan voor 1-goed antwoord danwel 0-fout antwoord. De waarschijnlijkheidsfunctie wordt dan als volgt:

$$L\,(u_1,u_2,\ldots\ldots,u_n \mid \theta) = \prod_{j=1}^{n} P_j^{u_i}\, Q_j^{1-u_i}$$

waar $P_j = P(U_j|\theta)$ en $Q_j = 1-P(U_j|\theta)$, de machtsverheffing in de formule zorgt ervoor dat in geval van een goed antwoord (1) P_i in het product meegenomen wordt, terwijl in geval van een fout antwoord (0) Q_i wordt meegenomen.

De waarde van θ die de waarschijnlijkheidsfunctie maximaliseert wordt aangeduid als de maximale waarschijnlijkheidsschatting van θ voor een kandidaat. Het vinden van deze schattingen op bovenstaande manier gebeurt door middel van een zoekproces door een computerprogramma. Efficiënte procedures gebruiken het feit dat, op het punt waar de functie zijn maximum bereikt, de helling van de functie (de eerste afgeleide) nul is. Dus de maximale waarschijnlijkheidsschatting kan gevonden worden door de vergelijking die verkregen wordt door de eerste afgeleide gelijk aan nul te stellen. Ook hier kan de vergelijking niet direct opgelost worden en moeten schattingsmethoden gebruikt worden. De meest gebruikte vorm van een dergelijke schattingsmethode is de Newton-Raphson procedure. Dit is een herhaaldelijk zoekproces waarin de parameterschattingen verbeterd worden (Hambleton, Swaminathan & Rogers,1991).

Tot Slot

In dit hoofdstuk is beschreven hoe de item- en vaardigheisparameters geschat kunnen worden voor dichotome modellen. De meest gebruikte methoden: de ‘joint-maximum-likelihood’ (JML), de ‘marginal-maximum-likelihood’ (MML) en de ‘conditional-maximum-likelihood’(CML) zijn beschreven. Dit hoofdstuk beperkte zich tot de schattingen bij de dichotome modellen. In het onderzoek is -met gebruik van het programma MULTILOG- gebruik gemaakt van een polytoom model en daarbij behorende schattingen. De behandeling van de schattingsmethode voor polytome modellen valt buiten het bereik van deze scriptie; de methode heeft de schattingsmethode voor dichotome modellen echter als grondslag.

Voordat de resultaten van het onderzoek besproken worden volgt nu eerst het hoofdstuk over de respondenten en analyses.

Methoden

De respondenten

De onderzochte vragenlijst bestaat uit 10 verschillende constructen, 5 van deze constucten zijn in dit onderzoek geanalyseerd. Het zijn de vragen die van toepassing zijn op de vaardigheden van de organisatie en de vaardigheden van de persoon zelf, commitment (naar de organisatie toe en naar het beroep toe), en persoon-organisatie fit. In de vragenlijst hebben deze constructen de volgende benamingen: 'vaardigheden van de organisatie', 'uw vaardigheden', 'uw houding ten opzichte van uw organisatie' (deze valt uiteen in 3 deelconstructen: affectief, normatief en continue), 'uw houding ten aanzien van uw beroep' (ook deze valt uiteen in dezelfde drie deelconstructen) en tot slot 'hoe u vindt dat de organisatie en u bij elkaar passen'.

De vragenlijst is door 1427 mensen ingevuld, uit 3 verschillende sectoren: een zorginstelling, een petrochemische fabriek en een ziekenhuis. Van deze 1427 hebben 1166 (81,7%) vrouwen en 256 (17,9%) mannen de vragenlijst ingevuld en 4 (0,3%) mensen hebben hun geslacht niet ingevuld. De respondenten van de zorginstelling zijn in 2 groepen verdeeld: het gewone personeel en de vrijwilligers. 597 mensen uit de zorginstelling hebben de vragenlijst ingevuld en 57 van de vrijwilligers. 550 mensen uit de petro chemische fabriek en 222 uit het ziekenhuis. 1 respondent heeft niet ingevuld waar hij/zij werkt.

In de tabellen hieronder staan de gegevens wat betreft de leeftijd van de respondenten, de tijd dat ze bij de organisatie werken en de opleiding die ze gevolgd hebben. Uit deze gegevens kunnen we opmaken dat het gaat om een redelijk goed verdeelde steekproef, alleen het percentage vrouwen is hoog te noemen. Dit is waarschijnlijk te verklaren omdat er in ziekenhuizen en zorginstellingen meer vrouwen dan mannen werkzaam zijn.

Leeftijd in jaren	Aantal	Procent
20	90	6,3
20-30	194	13,6
31-40	241	16,9
41-50	188	13,2
51-60	564	39,5
61+	137	9,6
onbekend	10	0,7

Werktijd in jaren	Aantal	Procent
< 1	90	6,3
1-2	194	13,6
2-5	241	16,9
6-10	188	13,2
11-15	564	39,5
15 +	137	9,6
onbekend	10	0,7

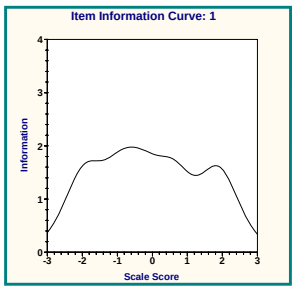
Opleiding	Aantal	Procent
Lager Onderwijs	20	1,4
Middelbaar Onderwijs	468	32,8
MBO	627	43,9
HBO	239	16,7
WO	27	1,9
Anders	4	0,3
Onbekend	41	2,8

Analyses

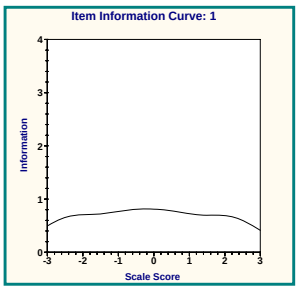
De gegevens zijn geanalyseerd met behulp van het programma MULTILOG. MULTILOG is een computerprogramma dat ontworpen is om item-analyse en bewerking van psychologische testen binnen het kader van Item Respons Theorie te vergemakkelijken. Zoals de naam al impliceert wordt MULTILOG gebruikt voor items met meerdere alternatieven en maakt het gebruik van logistische respons modellen, zoals het 1-parameter logistisch model, het 2-parameter logistisch model, het 3-parameter logistisch model, het Graded Respons Model, het Nominal Model en het Multiple-choice model.

De gegevens van het onderzoek zijn polytoom en hebben geordende responsen: volledig mee oneens – mee oneens – beetje mee oneens – niet eens/niet oneens – beetje mee eens – mee eens – volledig mee eens.
Om deze te analyseren is er dus een polytoom item respons model nodig voor geordende responsen. Het Partial Credit Model en het Graded Respons Model komen in aanmerking. Aangezien alleen het Graded Respons Model met behulp van het programma MULTILOG geanalyseerd kan worden, is voor dat model gekozen. De parameters zijn geschat middels de MML-schatting.

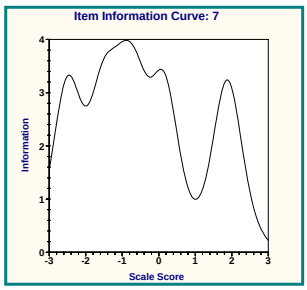
Om de grafieken en tabellen van de resultaten van het onderzoek beter te begrijpen worden deze hieronder in zijn algemeenheid toegelicht. Als eerste staan hier 3 voorbeelden van item informatie functies. Bij de eerste functie wordt de meeste informatie verkregen rond het gemiddelde deel van de latente vaardigheid (θ=0). Bij de tweede functie wordt er relatief weinig informatie gegeven, er wordt over het gehele vaardigheidsgebied ongeveer even precies gemeten. De derde functie geeft de meeste informatie: bij de lagere waarden van θ meet het item zeer precies, deze precisie wordt een stuk lager tussen de 0 en 1 en gaat vanaf een vaardigheid van 1 weer omhoog, waarna die bij een vaardigheid van 2 weer omlaag gaat.



Item informatie functie 1



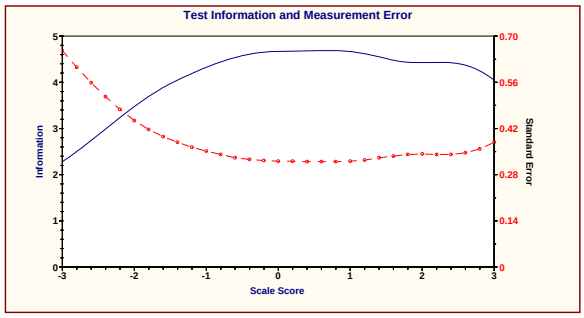
Item informatie functie 2



Item informatie functie 3

IRT modellen testen karakteristieken als functie van de respondent's positie op de theta; het betrouwbaarheidsbegrip in traditionele zin is dus niet betekenisvol in deze context: geen enkel getal kan de test-karakteristieken op alle theta-niveaus accuraat beschrijven. Echter, de zogenaamde *marginale* betrouwbaarheid-index probeert de gemiddelde betrouwbaarheid van de test te schatten over het theta-continuüm (Thissen,1986).

De doorlopende lijn in onderstaandestaande test informatie functie geeft de test informatie weer, de gestippelde lijn geeft de standaard meetfout weer. Over deze item informatie functie kun je zeggen dat hij vanaf vaardigheid –1 tot vaardigheid 3 ongeveer even goed meet. Bij een latente vaardigheid lager dan –1 meet de test iets minder precies en neemt de meetfout ook toe.
Duidelijk is ook te zien dat de beide functies elkaars spiegelbeeld zijn. Dit komt door de relatie tussen beide functies. De standaardmeetfout is gelijk aan: '1 gedeeld door de wortel uit de testinformatiefunctie'. Als voorbeeld uit onderstaand figuur nemen we de waarde van θ van –3. De testinformatie is 2,3 (linkerschaal) hieruit volgt dat de standaardmeetfout gelijk is aan $1 / \sqrt{2,3} = 0,66$ (zie rechterschaal). Dus hoe meer informatie je hebt, hoe groter de nauwkeurigheid is, en hoe kleiner de standaardfout.



De α-en β-waarden uit de tabel (zie blz. 36) dienen als volgt geïnterpreteerd te worden: de α-waarde is de dicriminatieparameter deze representeert de mate waarin het item onderscheid maakt tussen respondenten met verschillende θ-niveau's, de β-waarde geeft aan bij welke latente vaardigheid je de grootste kans hebt om in een bepaalde categorie te antwoorden. Neem als voorbeeld item 1: bij een latente vaardigheid lager dan –2,26 is de kans dat je in categorie 1 antwoordt het grootst, bij een latente vaardigheid tussen de –2,26 en -0,82 is de kans het grootst dat je in categorie 2 antwoordt, bij een latente vaardigheid tussen de –0,82 en –0,25 is de kans het grootst dat je in categorie 3 antwoordt, bij een latente vaardigheid tussen de –0,25 en 0,11 is de kans het grootst dat je in categorie 4 antwoordt, bij een latente vaardigheid tussen de 0,11 en 1,05 is de kans het grootst dat je in categorie 5 antwoordt en bij een latente vaardigheid tussen de 1,05 en 3,05 is de kans het grootste dat je antwoordt in categorie 6. Tot slot bij een latente vaardigheid van 3,05 of hoger is de kans het grootste dat je in categorie 7 antwoordt. Het geheel meet dus de latente vaardigheid van –2,26 tot 3,05.

De resultaten zijn ook geanalyseerd met behulp van de klassieke testtheorie (door middel van het computerprogramma SPSS). Er is aan de hand van Cronbach's Alpha gekeken naar de betrouwbaarheid van de test als geheel. Deze telt dus voor alle mensen ongeacht hun vaardigheid. Deze neemt aan dat de meetfout voor alle respondenten, ongeacht hun latente vaardigheid, hetzelfde is. Cronbach's Alpha wordt gebruikt bij testen met open vragen en wordt berekend uit de gemiddelde covariantie tussen de scores. Dit is een maat voor de variantie tussen de itemscores binnen het testsysteem. Het is de meest gebruikte schatting van interne consistentie van items in een schaal. Bij een alpha-waarde hoger dan 0,80 mag de test betrouwbaar genoemd worden. Is de totale alpha lager dan 0,60 en kan die niet verhoogd worden door items te hercoderen of te verwijderen, dan is het niet verantwoord om de schaal te handhaven. De items meten dan waarschijnlijk meer dan één begrip.

$$\alpha = \frac{N}{N-1}\left(1 - \frac{\Sigma S^2 i}{\Sigma S^2 x}\right)$$

Waarin N het aantal items is en $\Sigma S^2 i / \Sigma S^2 x$ de som van alle itemvarianties / de toetsvariantie.

Tot slot wordt er van elke schaal nog gekeken naar de minst informatieve vraag. Is deze vraag echt niet informatief, of kunnen we er toch nog andere informatie uithalen? Door middel van een variantie analyse is er gekeken of er een significant (0,01) verschil in antwoordpatroon is tussen respondenten van verschillende leeftijd, geslacht, opleiding, werktijd in jaren of sector waar ze werkzaam zijn. De exacte uitkomsten van deze analyse zijn te vinden in de bijlage, evenals de verklaring van de legenda.

Resultaten van het onderzoek

Inleiding

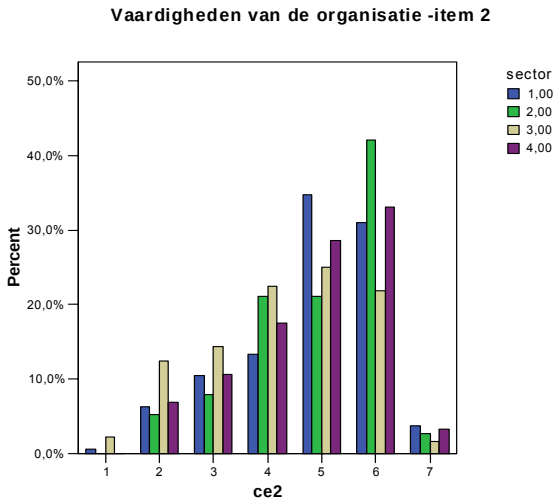
In dit hoofdstuk worden de resultaten van het onderzoek besproken. Per schaal wordt de betrouwbaarheid gemeten door middel van Cronbach's Alpha en wordt deze vergeleken met de betrouwbaarheid die verkregen is via de IRT analyse. De items die de minste en de items die de meeste informatie geven worden besproken en er wordt gekeken of de schaal betrouwbaarder kan worden gemaakt door weglating van het minst informatieve item. Vervolgens wordt het minst informatieve item nog geanalyseerd middels een variantie analye. Er wordt gekeken of er verschil in antwoordpatroon is tussen respondenten van verschillende leeftijd, geslacht, opleiding, werktijd in jaren of sector waar ze werkzaam zijn.
Aan het eind van elke paragraaf volgen dan de MULTILOG grafieken van de item karakteristieke curves, de item informatie cuves en de testinformatie curve en de tabel met de MULTILOG parameterschattingen voor de α en β waarden.

Vaardigheden van de organisatie

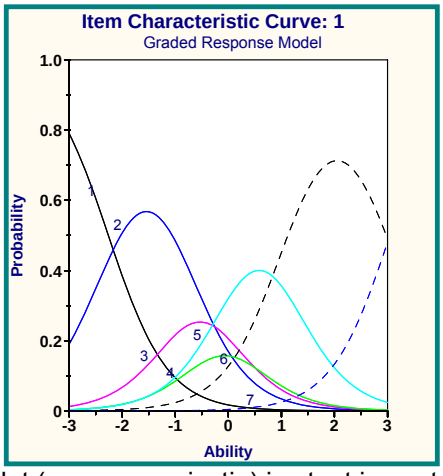
De 7 vragen in deze categorie gaan over hoe goed of hoe slecht de werknemer vindt dat de organisatie zijn taken of acties uitvoert.

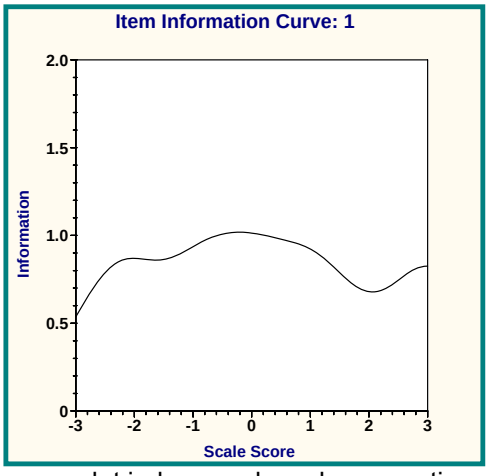
De betrouwbaarheid van deze schaal is volgens Cronbach's Alpha 0,832. De alpha score kan niet hoger worden door weglating van een item. Item 2 geeft weliswaar de minste informatie (dit item heeft de laagste item-totaal correlatie, zie ter illustratie de bijlage 'output SPSS'), maar de betrouwbaarheid van de test wordt dan 0,830 en wordt dus (iets) minder betrouwbaar. Ook als je naar de item informatie curve van item 2 ("Ik vind dat de organisatie in staat is om steeds nieuwe ideeën te ontwikkelen") kijkt, zie je dat deze bij de IRT analyse de minste informatie geeft. Item 4 "Ik vind dat de organisatie in staat is om hoge doelen te stellen en deze dan te bereiken" geeft de meeste informatie (α=2,21).
De test als geheel is dus –met een waarde groter dan 0,80 – betrouwbaar te noemen. Kijkend naar de β-waarden uit de IRT analyse valt op dat de schaal over het gehele deel van θ meet: item 1 van –2,26 tot 3,05; item 2 van –4,64 tot 3,77; item 3 van -2,77 tot 3,08; item 4 van –2,52 tot 2,72; item 5 van –4,12 tot 2,19; item 6 van –2,94 tot 2,82 en item 7 van –2,78 tot 2,43. Als er wordt gekeken naar het gemiddelde opgetelde antwoord in deze categorie dan is dat 31,11 (zie bijlage 'output SPSS'). Dat betekent dat alle antwoordpatronen (met een reeks van 1 tot en met 7) opgeteld over de 7 vragen uit deze categorie 31,11 opleveren. Het 'gemiddelde' antwoord op deze vragen is dan 4,44 (wat in de buurt van het verwachte gemiddelde van 4 ligt). Uit de testinformatiefunctie kunnen we nog aflezen dat de schaal respondenten meet met verschillende meningen over de vaardigheden van de organisatie, van laag tot hoog met bijna dezelfde mate van precisie. Alleen bij respondenten die slecht over de vaardigheden van de organisatie denken (met een θ van minder dan –2) is deze iets minder precies en bij respondenten met een θ die groter is dan 1, neemt de precisie van de schaal iets af en gaat de standaardmeetfout iets omhoog.

Het minst informatieve item 2 ‘Ik vind dat de organisatie in staat is om steeds nieuwe ideeën te ontwikkelen’ bekijken we door middel van een variantie analyse nog eens nader. We zien dan dat er geen significant verschil in antwoordpatroon is tussen respondenten van verschillend geslacht, opleiding, leeftijd en werktijd in jaren. Wel is er een verschil in antwoordpatroon tussen de respondenten uit de verschillende sectoren, dit is in het onderstaande histogram te zien. We zien dat vooral werknemers uit de zorginstelling (sector 1), vrijwilligers uit de zorginstelling (sector 2) en werknemers uit het ziekenhuis (sector 4) het met deze stelling eens zijn. Terwijl de werknemers van de petro-chemische fabriek (sector 3) deze vraag meer negatief beantwoord hebben.

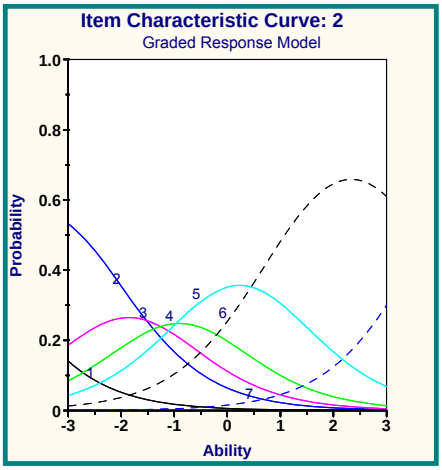


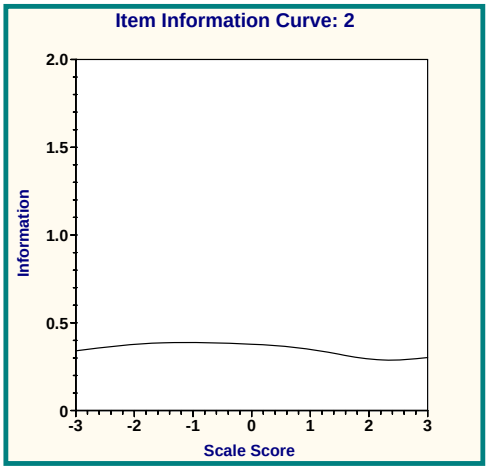
MULTILOG grafieken voor ‘vaardigheden van de organisatie’



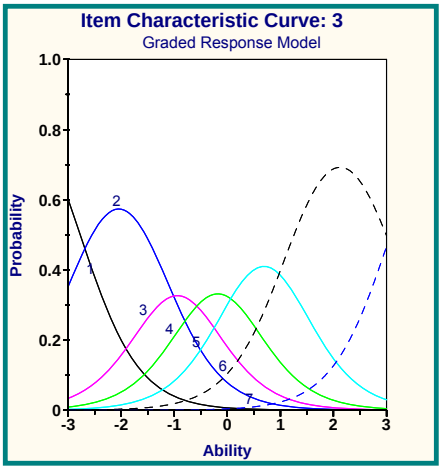


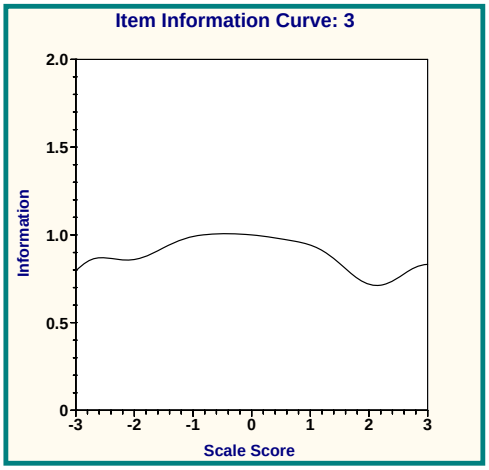
Ik vind dat (naam organisatie) in staat is om te zorgen dat iedere medewerker gemotiveerd is om zijn/haar werk te doen.



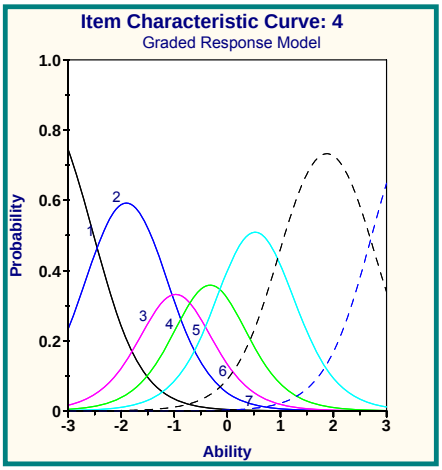


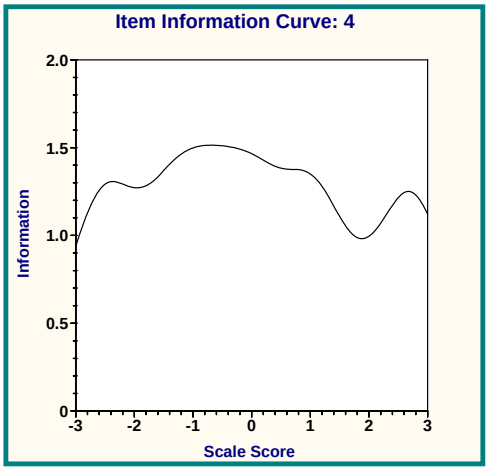
Ik vind dat (naam organisatie) in staat is om steeds nieuwe ideeën te ontwikkelen.



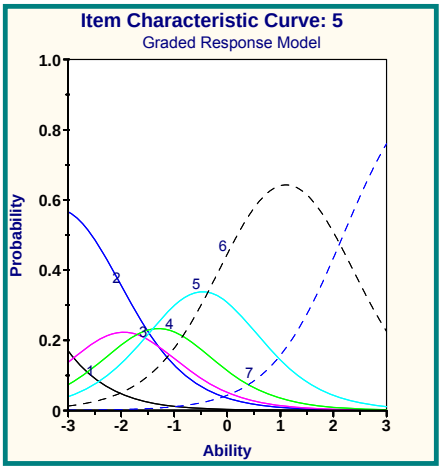


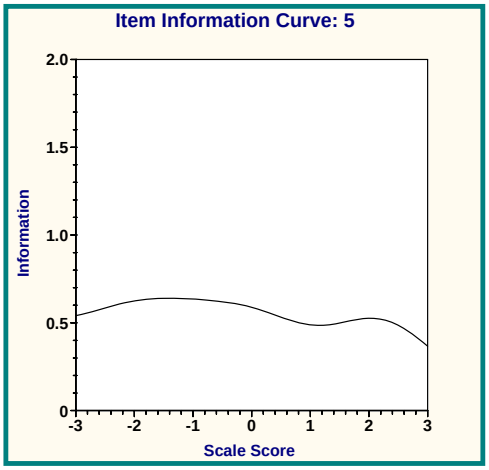
Ik vind dat (naam organisatie) in staat is om zich heel flexibel op te stellen ten aanzien van de omgeving.



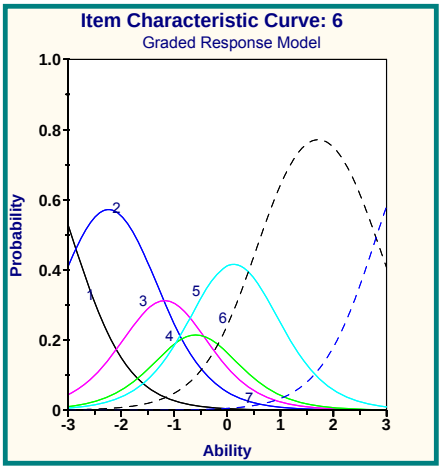


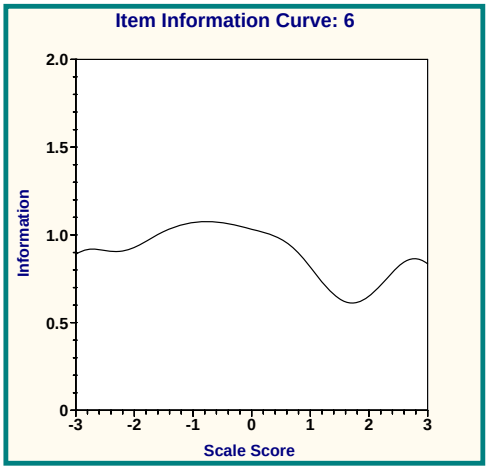
Ik vind dat (naam organisatie) in staat is om hoge doelen te stellen en deze dan te bereiken.



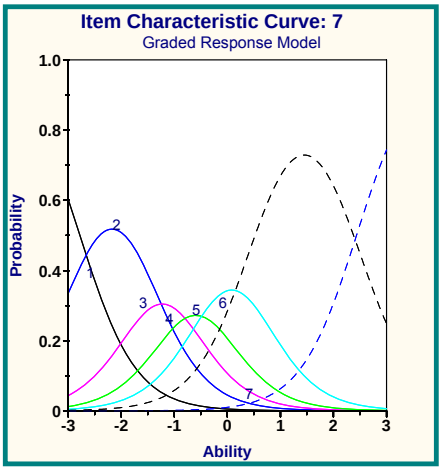


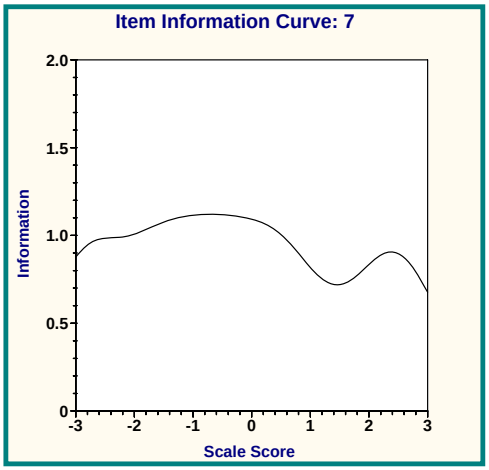
Ik vind dat (naam organisatie) in staat is om producten te leveren van zeer hoge kwaliteit.



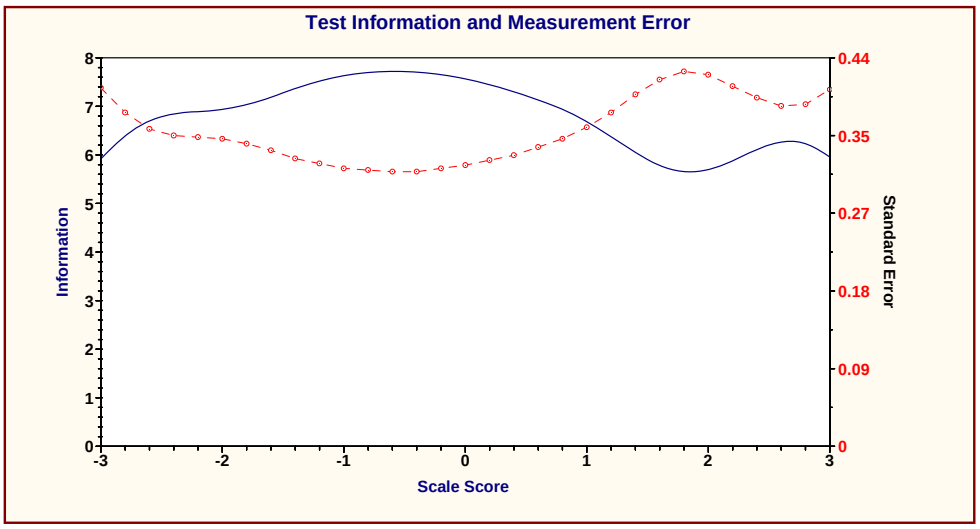


Ik vind dat (naam organisatie) in staat is om duidelijke informatie te verschaffen over het werk dat gedaan moet worden.





Ik vind dat (naam organisatie) in staat is om altijd betrouwbaar te zijn voor haar doelgroepen.



Marginale betrouwbaarheid van de test: 0,8604

MULTILOG parameter schattingen voor 'Vaardigheden van de organisatie'

	α (S.E.)	β1 (S.E.)	β2 (S.E.)	β3 (S.E.)	β4 (S.E.)	β5 (S.E.)	β6 (S.E.)
Item 1	1,79 (0,09)	-2,26 (0,12)	-0,82 (0,06)	-0,25 (0,05)	0,11 (0,05)	1,05 (0,07)	3,05 (0,19)
Item 2	1,10 (0,05)	-4,64 (0,41)	-2,35 (0,16)	-1,37 (0,10)	-0,45 (0,08)	0,90 (0,09)	3,77 (0,30)
Item 3	1,79 (0,07)	-2,77 (0,16)	-1,32 (0,07)	-0,56 (0,05)	0,20 (0,05)	1,18 (0,07)	3,08 (0,19)
Item 4	2,21 (0,09)	-2,52 (0,13)	-1,29 (0,06)	-0,66 (0,05)	0,02 (0,04)	1,03 (0,06)	2,72 (0,15)
Item 5	1,42 (0,08)	-4,12 (0,35)	-2,28 (0,13)	-1,64 (0,10)	-0,96 (0,07)	0,03 (0,06)	2,19 (0,14)
Item 6	1,84 (0,07)	-2,94 (0,18)	-1,53 (0,08)	-0,83 (0,05)	-0,36 (0,05)	0,60 (0,05)	2,82 (0,17)
Item 7	1,88 (0,11)	-2,78 (0,16)	-1,56 (0,08)	-0,89 (0,06)	-0,30 (0,05)	0,47 (0,05)	2,43 (0,13)

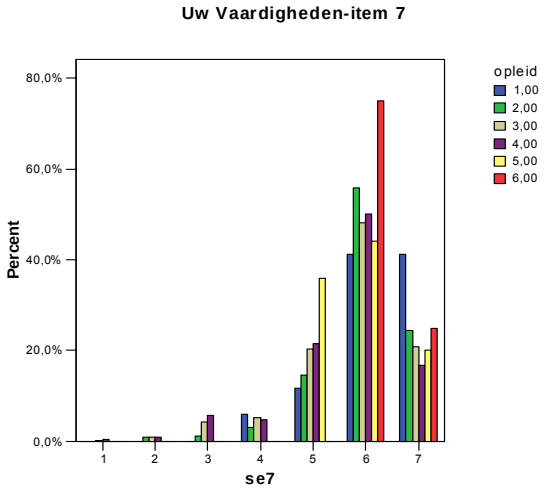
Uw Vaardigheden

De 7 vragen in deze categorie gaan over de vaardigheden die de respondenten, op het moment van invullen van de vragenlijst, hebben.

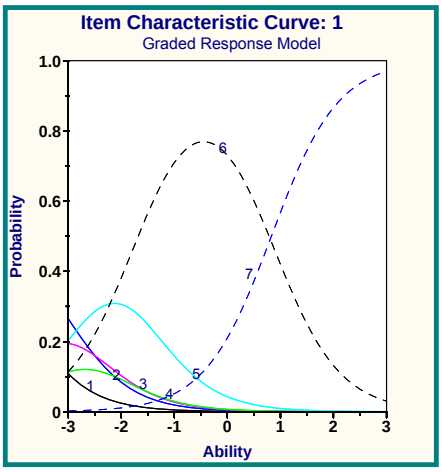
De betrouwbaarheid van deze schaal is, gemeten met Cronbach's Alpha 0,785. Net onder de kritieke waarde van 0,80. Maar ook door weglating van het minst

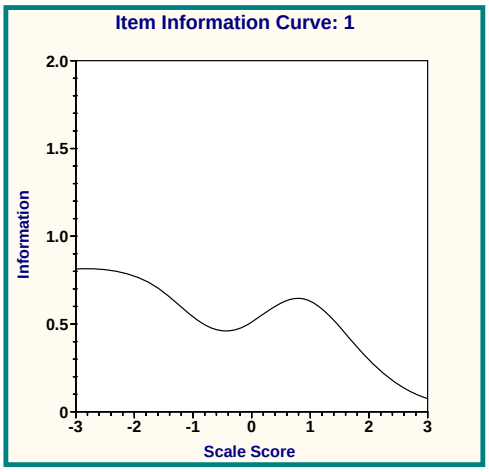
informatieve item -item 7- kan de betrouwbaarheid niet boven de 0,80 uitkomen. Deze wordt dan 0,781 en blijft dus nagenoeg gelijk. In de IRT analyse zie je ook dat item 7 “Ik ben in staat om heel ordelijk te werken” de minste informatie geeft (α=1,18). Als we kijken naar de β-waarden dan valt op dat de schaal voornamelijk meet voor een θ van ongeveer –4 tot +1 (item 1 van –4,32 tot 0,83; item 2 van –4,10 tot 1,22; item 3 van –4,42 tot 0,96; item 4 van –4,47 tot 1,83; item 5 van –3,66 tot 1,24; item 6 van –5,01 tot 0,96 en item 7 van –6,12 tot 1,33). Dit is ook goed te zien in de test informatiecurve: vanaf latente vaardigheid 1 daalt de informatiecurve en stijgt de standaardmeetfout-curve. Als we kijken naar de opgetelde gemiddelde score van de respondenten dan is deze 40,88 wat een gemiddelde score per vraag oplevert van 5,84. De meeste vragen zijn dus beantwoord met score 5, 6 of 7 wat respectievelijk ‘beetje mee eens’, ‘mee eens’ en ‘volledig mee eens’ inhoudt. Als je naar de item informatie curves kijkt van deze schaal zie je ook dat mensen met een latente vaardigheid van 1 of hoger voornamelijk ‘volledig mee eens’ als antwoord hebben gegeven. Mensen met een hoge vaardigheid zijn zich dus heel goed bewust van hun eigen capaciteiten.
Item 5 “Ik ben in staat werk af te leveren van zeer hoge kwaliteit” geeft de meeste informatie (α=2,40). Dit item geeft vooral veel informatie in het latente vaardigheid gebied beneden de -1, dit is goed te zien aan zowel de item informatiecurve als de item karakteristieke curve. Bij een latente vaardigheid van 0 zakt de item informatie curve, dit komt omdat de kans dat je daar antwoord 6 (mee eens) geeft heel hoog is, ook na een latente vaardigheid van 1 zakt de curve weer, dit keer omdat de kans om antwoord 7 te geven (volledig mee eens) daar heel hoog is.

Het minst informatieve item, item 7 ‘Ik ben in staat om heel ordelijk te werken’ is verder bekeken door middel van een variantie analyse. Hieruit blijkt dat er geen verschil is tussen de antwoordpatronen van mannen en vrouwen en dat er kleine verschillen zijn tussen respondenten met verschillende opleiding, leeftijd, werktijd in jaren en sector waar ze werkzaam zijn. Het grootste verschil in antwoordpatroon is te vinden in respondenten van verschillend opleidingsniveau, dit is te zien in onderstaand histogram. We zien hierin dat hoogopgeleide mensen alleen maar positief op de vraag hebben geantwoord en dat de negatieve antwoorden komen van mensen met een lagere opleiding. Wat dan wel weer opvallend is, is dat het antwoord ‘volledig mee eens’ (7) dan weer wel veel vaker voor komt bij de laagopgeleiden.

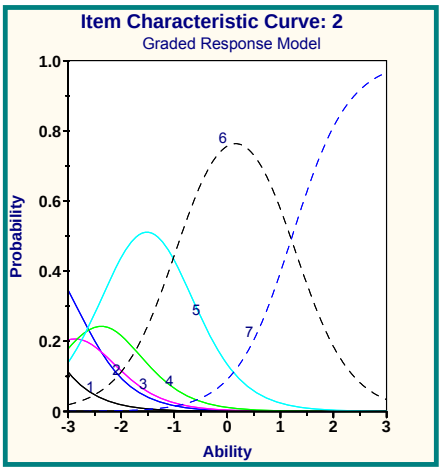


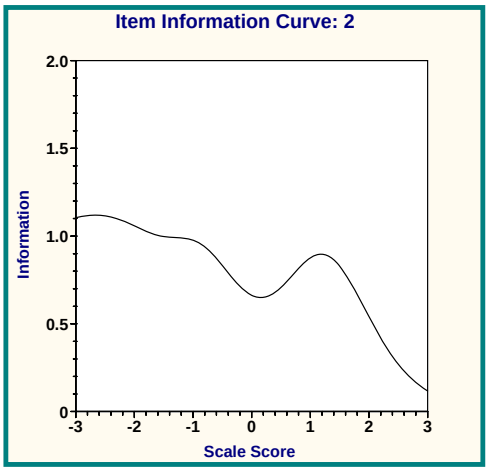
MULTILOG grafieken voor 'uw vaardigheden'



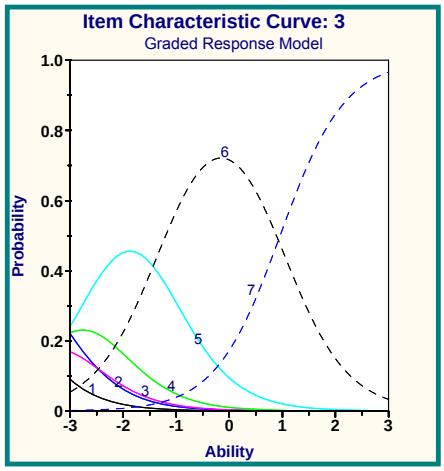


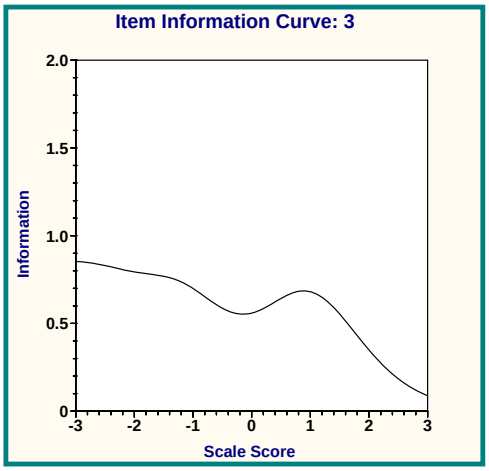
Ik ben in staat om binnen een team intensief samen te werken.



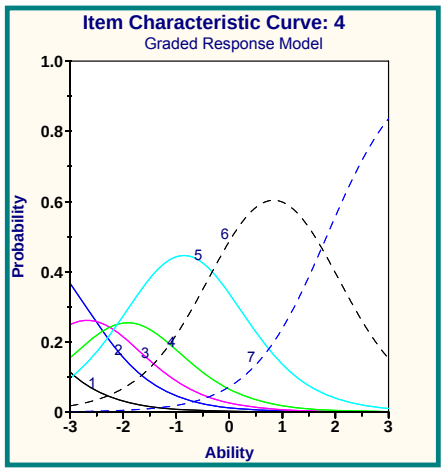


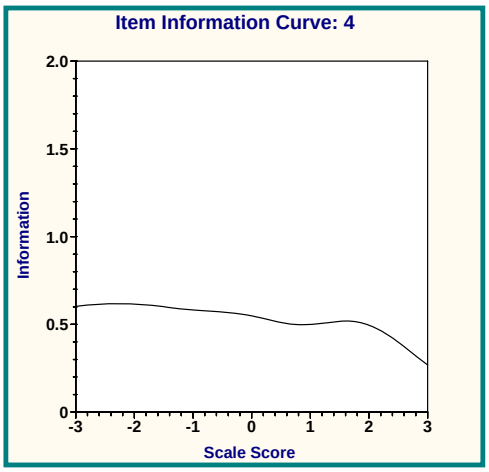
Ik ben in staat om problemen creatief op te lossen en nieuwe ideeën te bedenken.



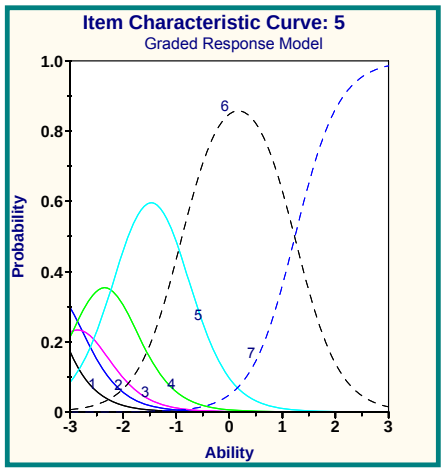


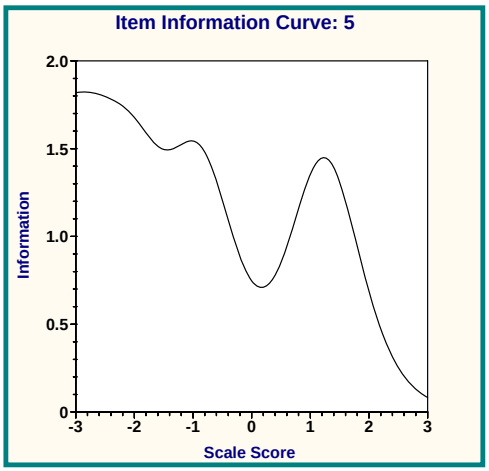
Ik ben in staat om me heel flexibel op te stellen wanneer de situatie daar onverwacht om vraagt.



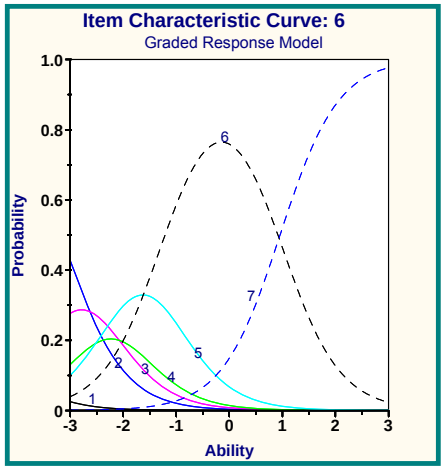


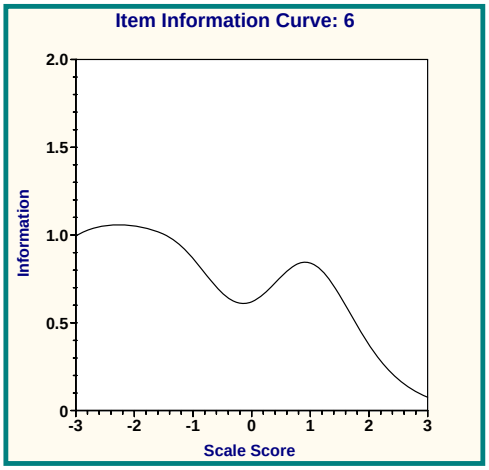
Ik ben in staat om me helemaal vast te bijten in gestelde doelen teneinde deze te bereiken.



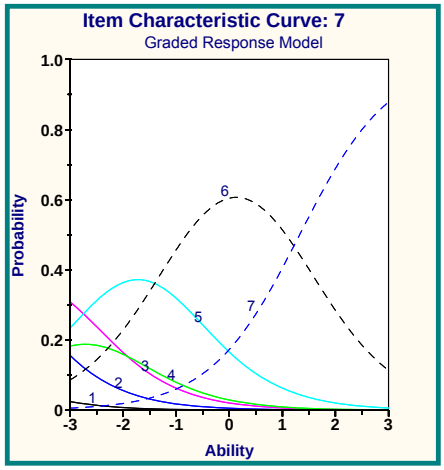


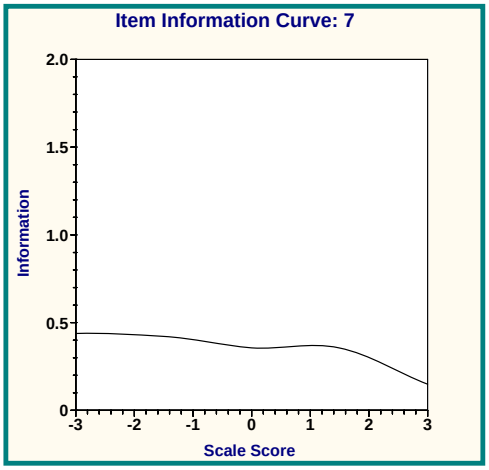
Ik ben in staat om werk af te leveren van zeer hoge kwaliteit.



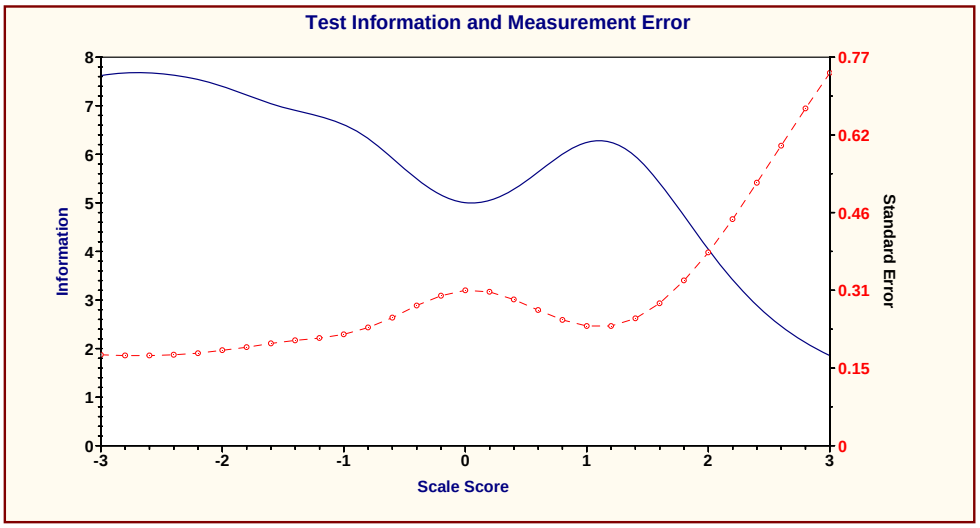


Ik ben in staat om alle verantwoordelijkheden die ik krijg te dragen.





Ik ben in staat om heel ordelijk te werken.



Marginale betrouwbaarheid van de test: 0,8223.

MULTILOG parameterschattingen van 'Uw vaardigheden'

	α (S.E.)	β1 (S.E.)	β2 (S.E.)	β3 (S.E.)	β4 (S.E.)	β5 (S.E.)	β6 (S.E.)
Item 1	1,59 (0,09)	-4,32 (0,47)	-3,33 (0,25)	-2,83 (0,19)	-2,53 (0,16)	-1,73 (0,10)	0,83 (0,07)
Item 2	1,88 (0,10)	-4,10 (0,45)	-3,09 (0,21)	-2,64 (0,16)	-2,12 (0,12)	-0,92 (0,06)	1,22 (0,07)
Item 3	1,63 (0,09)	-4,42 (0,51)	-3,49 (0,27)	-3,05 (0,21)	-2,47 (0,15)	-1,27 (0,08)	0,96 (0,07)
Item 4	1,39 (0,08)	-4,47 (0,44)	-3,06 (0,21)	-2,29 (0,14)	-1,54 (0,10)	-0,17 (0,06)	1,83 (0,11)
Item 5	2,40 (0,13)	-3,66 (0,39)	-3,06 (0,22)	-2,66 (0,16)	-2,04 (0,10)	-0,90 (0,05)	1,24 (0,06)
Item 6	1,82 (0,10)	-5,01 (0,96)	-3,11 (0,22)	-2,46 (0,14)	-2,01 (0,11)	-1,25 (0,07)	0,96 (0,06)
Item 7	1,18 (0,08)	-6,12 (0,90)	-4,28 (0,36)	-3,04 (0,22)	-2,39 (0,17)	-1,06 (0,09)	1,33 (0,10)

Uw houding ten opzichte van uw organisatie

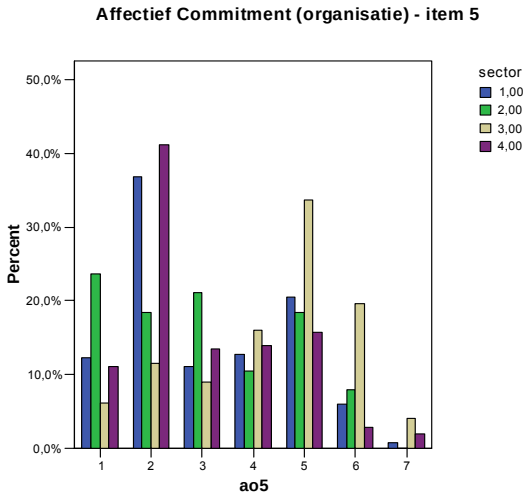
Affectief commitment

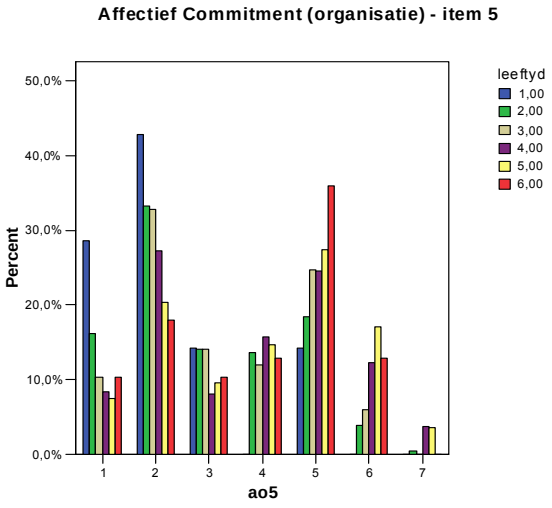
De 6 vragen uit deze schaal gaan over de mate waarin een werknemer zich betrokken voelt bij de organisatie. Het gaat hier om affectief commitment naar de organisatie toe.

De betrouwbaarheid, gemeten met Cronbach's Alpha, is voor deze schaal 0,842. Deze kan zelfs nog verhoogd worden naar 0,867 als het minst informatieve item, nummer 5: "ik ervaar problemen van deze organisatie als mijn eigen problemen", wordt weggelaten (α=0,82). In de IRT analyse zie je bij de item informatie curve van item 5 ook heel goed dat deze nauwelijks informatie geeft, de lijn loopt praktisch horizontaal en loopt bijna gelijk aan de x-as.
Uit de test informatie curve kunnen we aflezen dat de schaal voornamelijk goed meet op het gebied van een θ tussen de –2 en +1,5. Dit is ook af te lezen uit de β-waarden: item 1 –1,92 tot 1,88; item 2 –2,41 tot 1,62; item 3 –2,92 tot 1,40; item 4 –2,27 tot 1,64; item 5 –3,00 tot 4,49 en item 6 van –1,92 tot 2,45.
Item 4 "de organisatie betekent veel voor mij" geeft de meeste informatie (α=3,23). Opvallend in de item karakteristieke curve is dat de antwoordmogelijkheden 2,3,4,5 en 6 heel gepiekte smalle curves hebben. Zo is de kans dat respondenten met een θ van rond de –2 (respondenten met een negatieve houding naar de organisatie toe) voornamelijk 'mee oneens'(2) antwoorden, respondenten met een θ van rond de –1

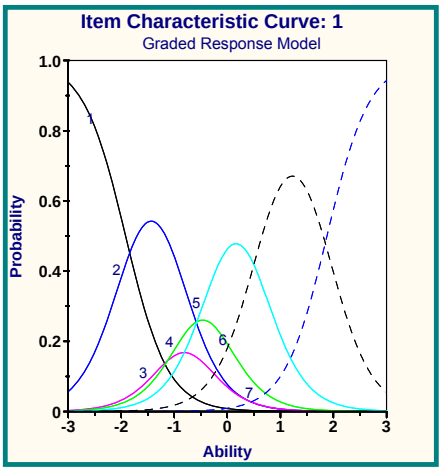
‘beetje mee oneens’(3) antwoorden, respondenten met een θ van rond de –0,5 ‘niet eens/niet oneens’(4) antwoorden, respondenten met een θ van rond de 0 ‘beetje mee eens’(5) antwoorden en respondenten met een θ van rond de 1 (respondenten met een redelijk positieve houding naar de organisatie toe) ‘mee eens’(6) antwoorden, heel groot.
De gemiddelde opgetelde score van deze schaal is 27,03 wat een gemiddeld antwoord geeft van 4,5. Dit is niet een extreem gemiddelde, en betekent dat alle antwoorden in een normale verdeling voorkomen.

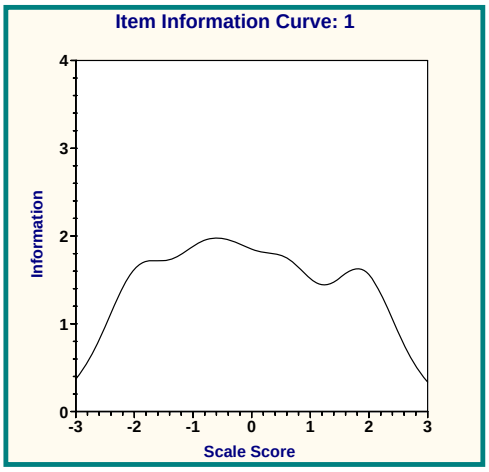
Het minst informatieve item, nummer 5 ‘ik ervaar problemen van deze organisatie als mijn eigen problemen’ zorgt voor een hogere betrouwbaarheid als deze wordt weggelaten. Door middel van de variantie analyse zien we dat er geen verschil is tussen de antwoorden van mannen en vrouwen. Voor de andere 4 groepen (opleiding, leeftijd, sector en aantal werkjaren) is er wel verschil. Het grootste verschil zit in de antwoorden van de respondenten uit de verschillende sectoren (1=zorginstelling personeel; 2=vrijwilligers zorginstelling; 3= personeel uit de petro chemische industrie; 4=ziekenhuispersoneel) en de leeftijden, dit is te zien in onderstaande histogrammen. Wat hier erg opvalt is dat het zorgpersoneel (1,2 en 4) problemen van de organisatie niet als hun eigen problemen zien en de respondenten uit de petro chemische industrie dit juist wel veelvuldig geantwoord hebben. Verder blijkt dat jonge werknemers (1=jonger dan 20 tot 6=61 of ouder) ook vaker negatief geantwoord hebben, terwijl oudere werknemers weer vaker aangeven de problemen van de organisatie juist wel als hun eigen problemen te zien.



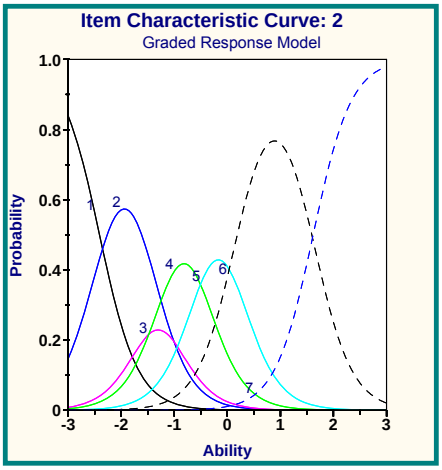


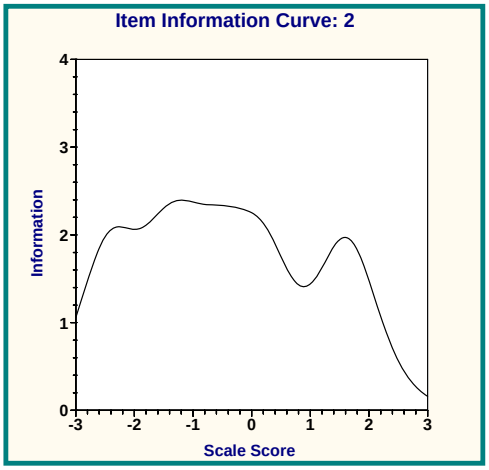
MULTILOG grafieken voor 'Affectief Commitment (organisatie)'



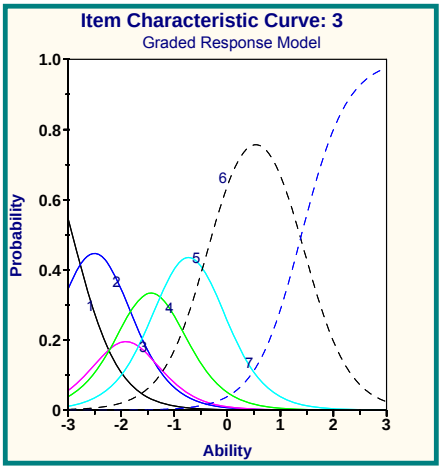


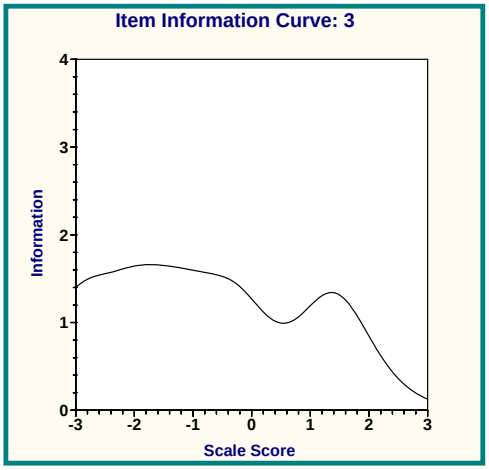
Ik voel me emotioneel gehecht aan (naam organisatie).



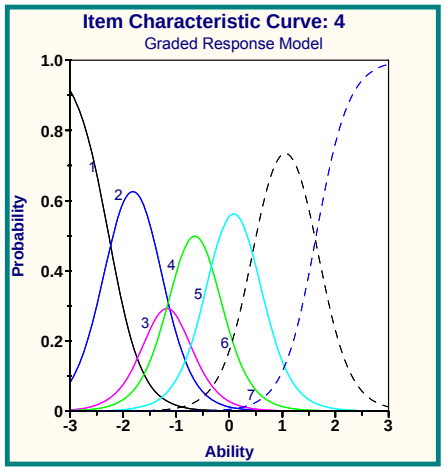


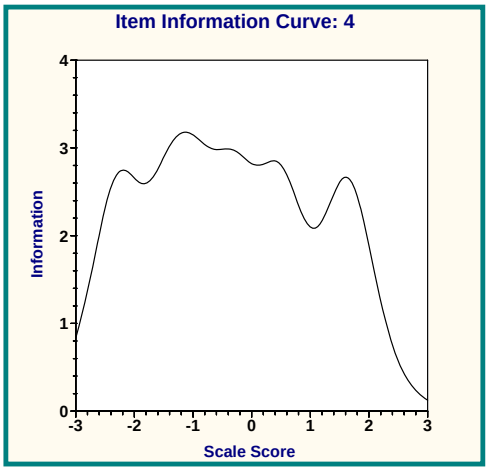
Ik ben er trots op aan anderen te vertellen dat ik bij (naam organisatie) werk.



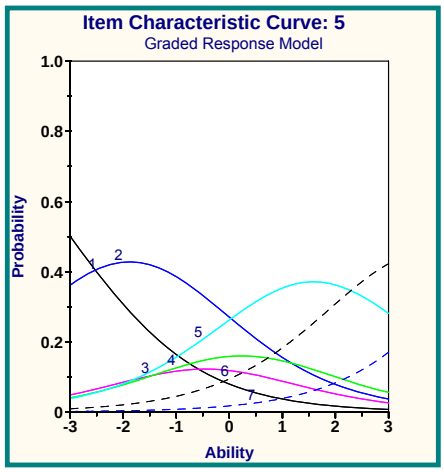


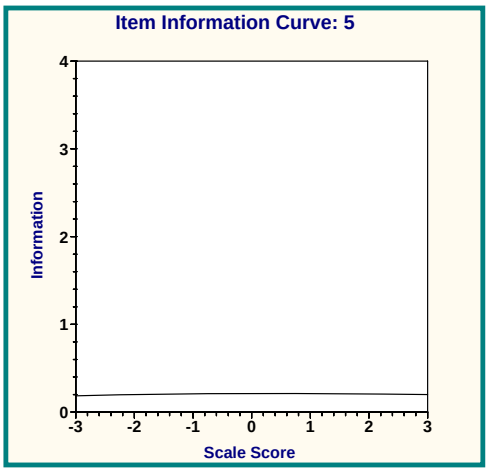
Ik voel me thuis bij (naam organisatie).



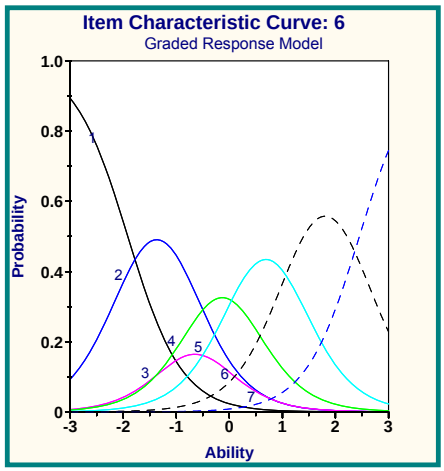


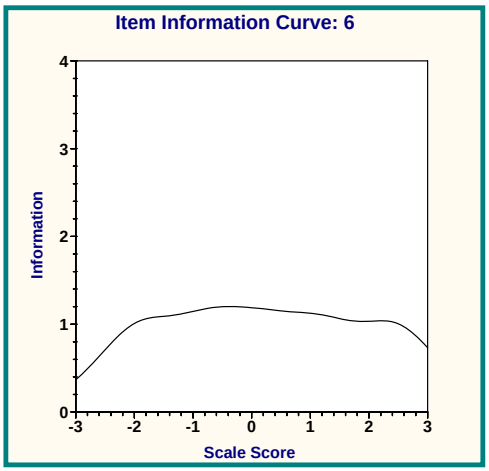
(naam organisatie) betekent veel voor mij.



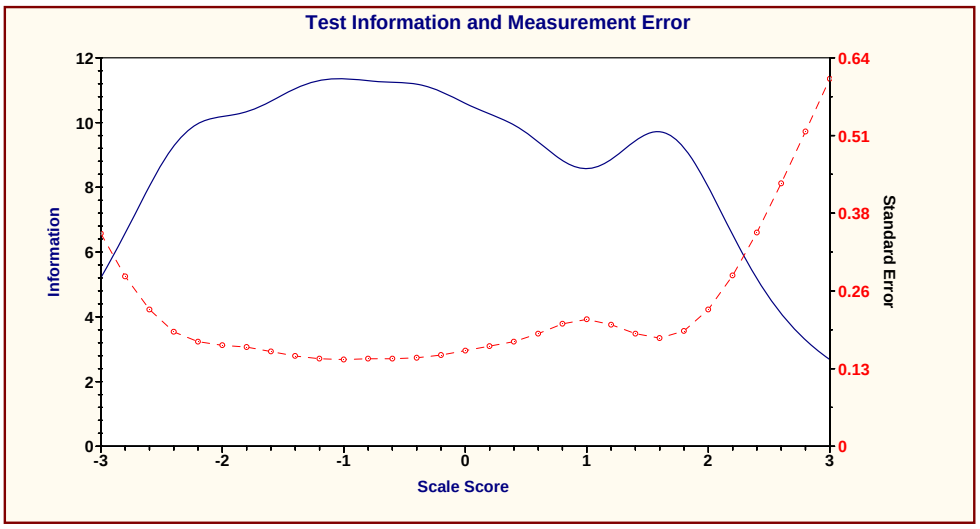


Ik ervaar problemen van deze organisatie als mijn eigen problemen.





Ik voel me als “een deel van de familie” in deze organisatie.



Marginale betrouwbaarheid van de test: 0,8998

MULTILOG parameterschattingen van 'Affectief Commitment (organisatie)'

	α (S.E.)	β1 (S.E.)	β2 (S.E.)	β3 (S.E.)	β4 (S.E.)	β5 (S.E.)	β6 (S.E.)
Item 1	2,50 (0,10)	-1,92 (0,09)	-0,95 (0,05)	-0,68 (0,04)	-0,25 (0,04)	0,58 (0,04)	1,88 (0,08)
Item 2	2,78 (0,11)	-2,41 (0,13)	-1,47 (0,06)	-1,14 (0,05)	-0,50 (0,04)	0,16 (0,04)	1,62 (0,07)
Item 3	2,29 (0,11)	-2,92 (0,20)	-2,09 (0,10)	-1,74 (0,08)	-1,14 (0,06)	-0,32 (0,04)	1,40 (0,07)
Item 4	3,23 (0,13)	-2,27 (0,11)	-1,36 (0,05)	-0,99 (0,04)	-0,31 (0,03)	0,47 (0,04)	1,64 (0,06)
Item 5	0,82 (0,06)	-3.00 (0,28)	-0,76 (0,11)	-0,16 (0,10)	0,63 (0,10)	2,54 (0,21)	4,49 (0,43)
Item 6	1,95 (0,09)	-1,92 (0,10)	-0,82 (0,06)	-0,48 (0,05)	0,21 (0,05)	1.16 (0,06)	2,45 (0,13)

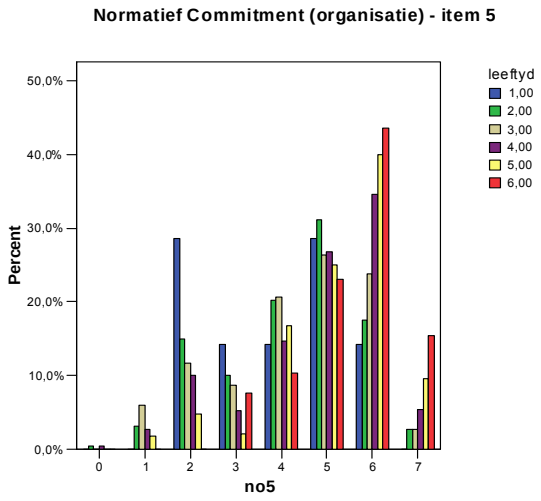
Normatief commitment

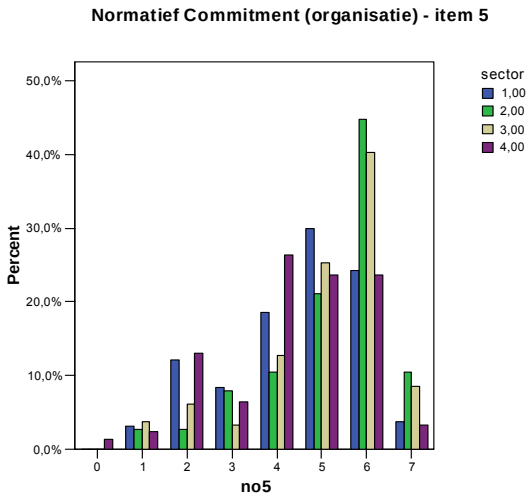
Deze 5 vragen gaan over de mate waarin een werknemer zich betrokken voelt bij de organisatie. Het gaat hier om normatief commitment naar de organisatie toe.

De betrouwbaarheid van deze schaal is 0,727 gemeten met Cronbach's alpha. Uit de analyse blijkt niet dat het ene item beduidend slechter is dan het andere, en de betrouwbaarheid kan dus ook niet verhoogd worden door weglating van een item. Uit de item informatie curve van de IRT analyse is te zien dat item 5 (α=1,02) "Ik vind dat iemand loyaal zou moeten zijn ten opzichte van zijn of haar organisatie" het minst

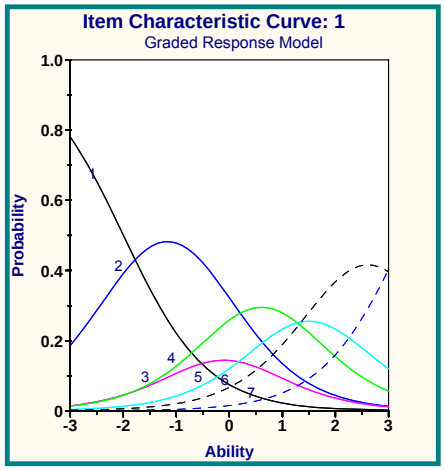
informatieve item is. Maar ook item 1 (α=1,26) “Het zou een goede zaak zijn als werknemers het grootste deel van hun loopbaan bij één organisatie zouden blijven” geeft weinig informatie. Opvallend is dat deze 2 vragen gaan over een waarde-oordeel over anderen (evenals vraag 3, die dan weer wel informatief is met een α van 1,73). De vragen 2 “Een van de belangrijkste redenen waarom ik bij de organisatie blijf, is dat ik loyaliteit belangrijk vind” (α=1,63) en vooral vraag 4 “Ik ben opgegroeid met de gedacht dat het waardevol is loyaal te blijven aan een organisatie” (α=1,93) geven meer informatie.
Kijkend naar de testinformatie curve kunnen we zeggen dat de schaal vanaf een θ van –1,5 op eenzelfde niveau van precisie meet. Dit is ook weer goed te zien aan de β-waarden uit de parameterschatting tabel.
De gemiddelde opgetelde score van deze schaal is 17,80 dit geeft een gemiddelde antwoord van 3,56. Een enigszins lage gemiddelde score die suggereert dat werknemers loyaliteit (het normatieve commitment) naar hun werkgever niet zeer belangrijk achten.

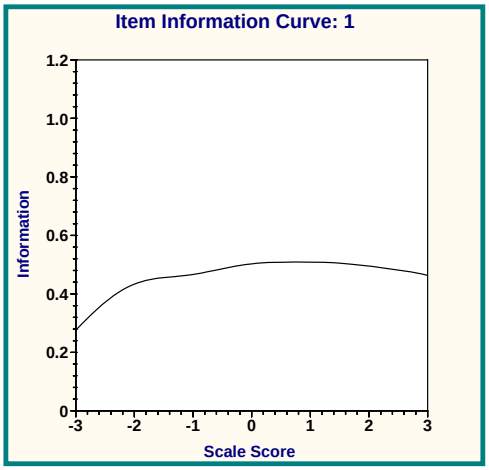
Item 5 ‘Ik vind dat iemand loyaal zou moeten zijn ten opzichte van zijn of haar organisatie’ geeft de minste informatie uit deze schaal. Uit de variantie analyse blijkt dat er geen significant verschil is tussen respondenten van verschillende opleiding, er een klein verschil is tussen respondenten van verschillend geslacht en het aantal dienstjaren en er een groot verschil is tussen respondenten met verschillende leeftijd en de verschillende sectoren. Deze laatste twee zijn opgenomen in onderstaande histogrammen. Hieruit blijkt dat jonge mensen het minder nemen met de loyaliteit naar de organisatie toe dan oudere mensen. Ook de vrijwilligers uit de zorginstelling (2) en de werknemers van de petrochemische industrie (3) hechten meer waarde aan de loyaliteit naar de organisatie toe dan het ziekenhuis- en zorginstelling personeel.



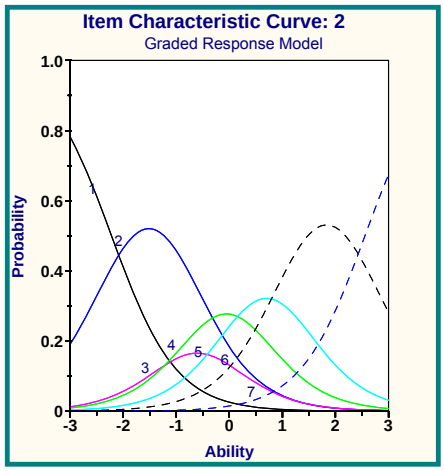


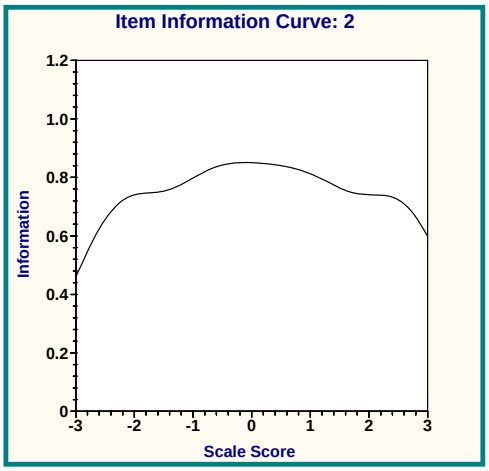
MULTILOG grafieken voor 'Normatief Commitment (organisatie)'



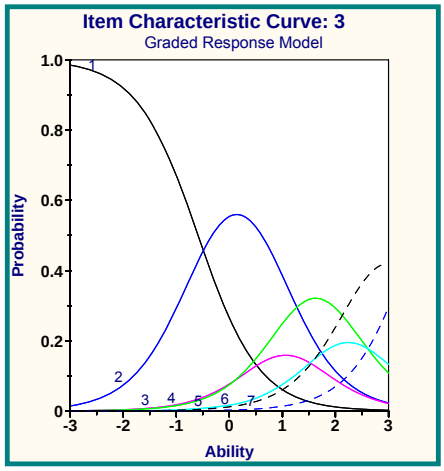


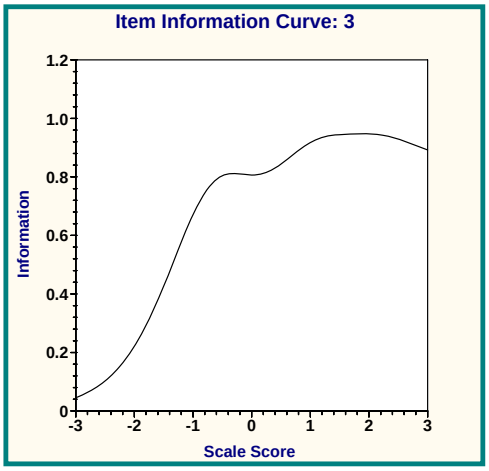
Het zou een goede zaak zijn als werknemers het grootste deel van hun loopbaan bij één organisatie zouden blijven.



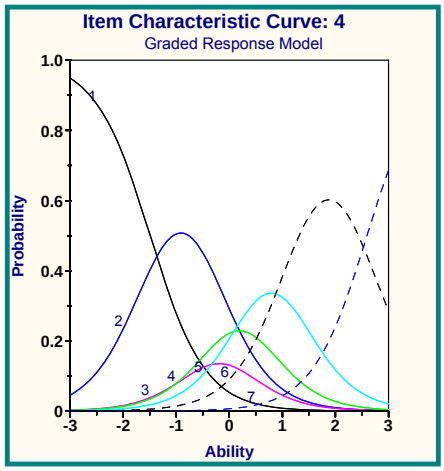


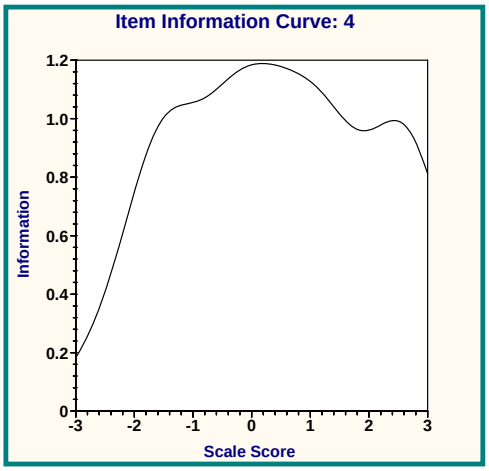
Een van de belangrijkste redenen waarom ik bij (naam organisatie) blijf, is dat ik loyaliteit belangrijk vind.



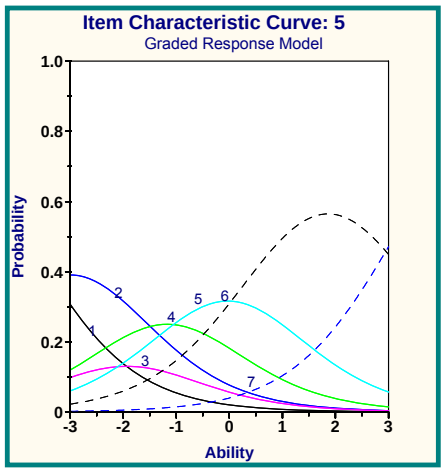


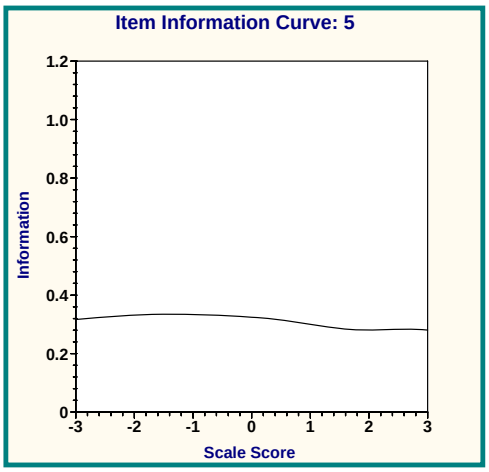
Het is onbehoorlijk om van de ene organisatie naar de andere over te stappen.



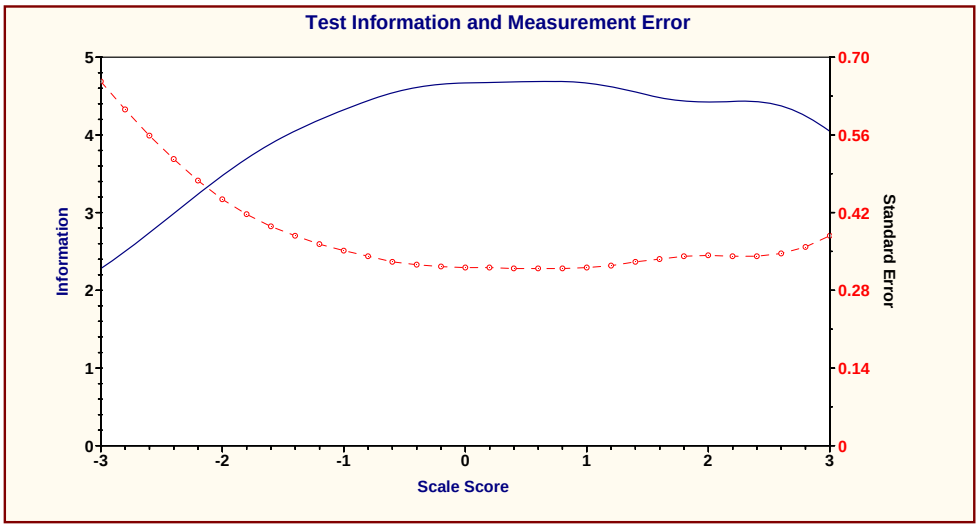


Ik ben opgegroeid met de gedachte dat het waardevol is loyaal te blijven aan een organisatie.





Ik vind dat iemand loyaal zou moeten zijn ten opzichte van zijn of haar organisatie.



Marginale betrouwbaarheid van de test: 0,7761

MULTILOG parameterschattingen van 'Normatief Commitment (organisatie)'

	α (S.E.)	β1 (S.E.)	β2 (S.E.)	β3 (S.E.)	β4 (S.E.)	β5 (S.E.)	β6 (S.E.)
Item 1	1,26 (0,07)	-2,00 (0,13)	-0,33 (0,06)	0,13 (0,06)	1,09 (0,08)	1,91 (0,12)	3,31 (0,23)
Item 2	1,63 (0,07)	-2,22 (0,12)	-0,81 (0,06)	-0,40 (0,05)	0,29 (0,05)	1,11 (0,07)	2,56 (0,15)
Item 3	1,73 (0,09)	-0,59 (0,06)	0,87 (0,06)	1,24 (0,07)	2,01 (0,11)	2,47 (0,14)	3,50 (0,22)
Item 4	1,93 (0,09)	-1,49 (0,08)	-0,33 (0,05)	-0,05 (0,04)	0,43 (0,05)	1,15 (0,06)	2,59 (0,15)
Item 5	1,02 (0,05)	-3,80 (0,28)	-2,18 (0,15)	-1,67 (0,12)	-0,67 (0,08)	0,61 (0,09)	3,11 (0,24)

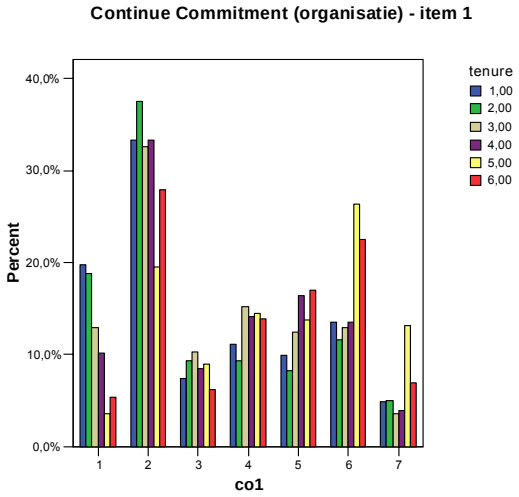
Continue commitment

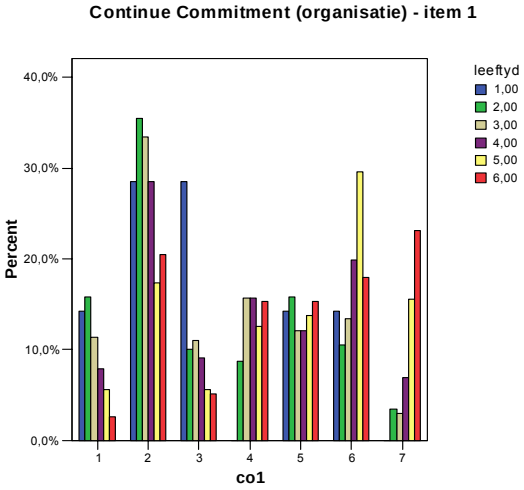
Deze 5 vragen gaan ook over de mate waarin een werknemer zich betrokken voelt bij de organisatie. Het gaat nu om continue commitment naar de organisatie toe.

Cronbach's alpha meet een betrouwbaarheid van 0,835 voor deze schaal, wat dus een goed resultaat te noemen is. Uit de analyse blijkt dat item 1 en item 5 het minst goed meten, maar door weglating van deze items gaat de betrouwbaarheid niet omhoog. De items die het beste meten zijn de nummers 2 en 3. Uit de IRT analyse blijkt ook dat de items 1 (α=1,52) "Het zou voor mij op dit moment moeilijk zijn om weg te gaan bij de organisatie, ook al zou ik dat willen" en 5 (α=1,61) "Ik ben bang voor wat er zou kunnen gebeuren als ik mijn baan opzeg, zonder meteen een nieuwe baan te hebben" weinig informatie geven; de item informatie curve loopt bijna vlak en bijna gelijk aan de x-as. Uit deze analyse blijkt ook dat item 4 (α=1,75) "Er zou teveel in mijn leven verstoord worden als ik nu ontslag zou nemen" een dergelijke item informatie curve heeft en dus ook weinig informatie verschaft. Dit in tegenstelling tot de items 2 (α=4,14) en 3 (α=4,08) (resp. "Als ik ontslag neem wordt het moeilijk om een andere baan te vinden" en "Ik heb het gevoel dat ik te weinig alternatieven heb om nu ontslag te nemen").
Als we naar de test informatie curve kijken zien we dat deze schaal vooral goed meet op het gebied tussen een θ van –1,5 en +1,5. Dit is ook goed af te lezen uit de parameterschatting tabel (voor item 1 van -1,96 tot 2,13; item 2 van –1,06 tot 1,48; item 3 van –1,16 tot 1,55; item 4 van –1,96 tot 1,76 en item 5 van –1,87 tot 1,65). Het gemiddelde van de opgetelde scores is 19,56; dat is een gemiddeld antwoord per vraag van 3,9.

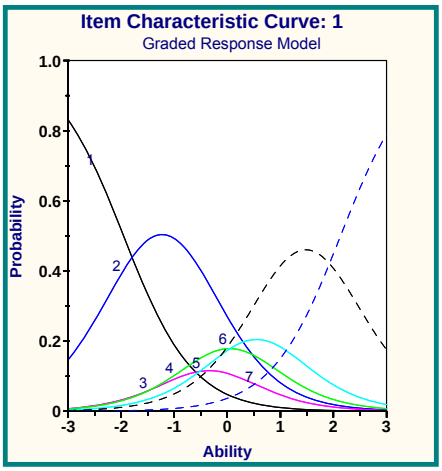
Item 1 'Het zou voor mij op dit moment moeilijk zijn om weg te gaan bij de organisatie, ook al zou ik dat willen' geeft de minste informatie. Middels de variantie

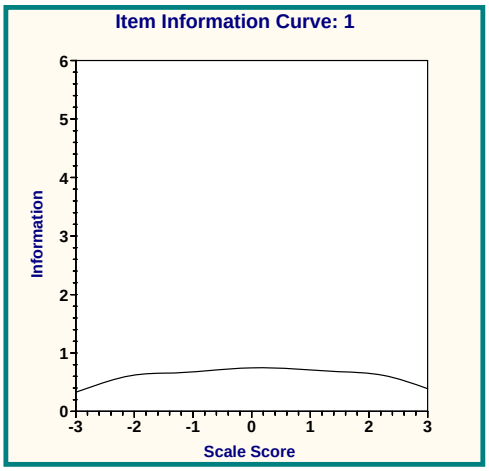
analyse zien we dat er geen significant verschil is in antwoordenpatroon tussen mannen en vrouwen, er verschil is tussen respondenten met een verschillende opleiding en sector en dat er een goed zichtbaar verschil is in antwoorden van respondenten met een verschillende leeftijd en het aantal jaren dat de respondenten bij de organisatie werken. Het eerste histogram laat zien dat respondenten die nog niet zo lang bij een organisatie werken (1=korter dan 1 jaar; 2=1 tot 2 jaar; 3=2 tot 5 jaar; 4=6 tot 10 jaar; 5=11 tot 15 jaar en 6=langer dan 15 jaar) minder moeite hebben om de organisatie te verlaten dan mensen die al langer bij de organisatie werken. Hetzelfde is te zien in het tweede histogram, die de leeftijd van de respondenten weergeeft. Hierin is te zien dat jongere mensen minder moeite hebben om de organisatie te verlaten dan oudere mensen.



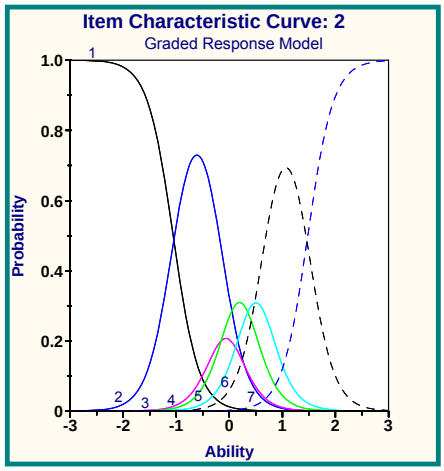


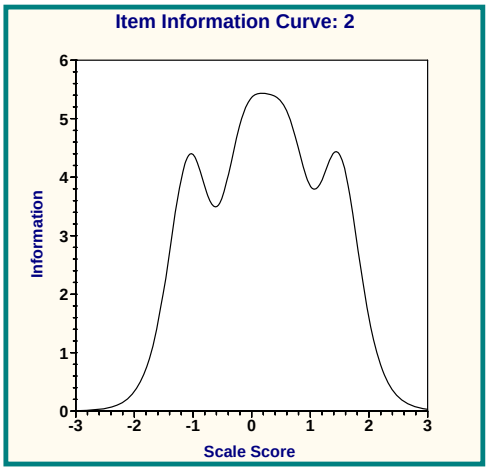
MULTILOG grafieken voor 'Continue Commitment (organisatie)'



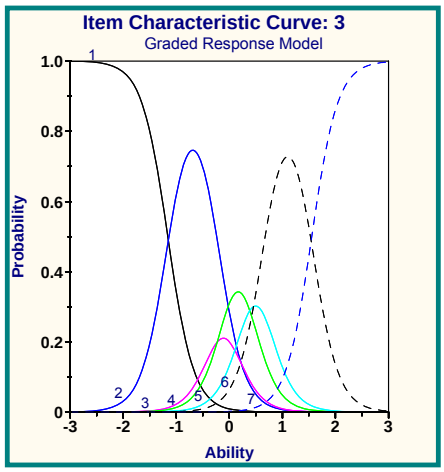


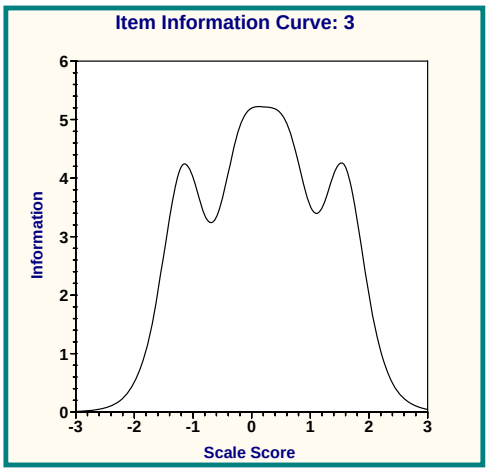
Het zou voor mij op dit moment moeilijk zijn om weg te gaan bij (naam organisatie), ook al zou ik dat willen.



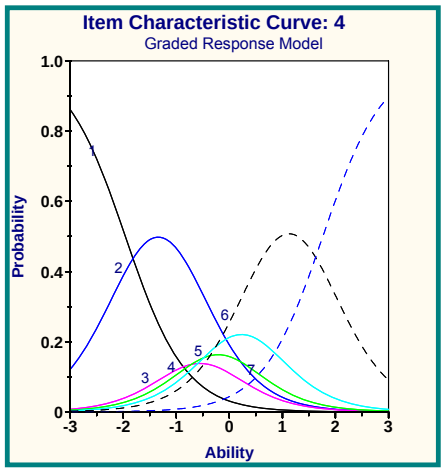


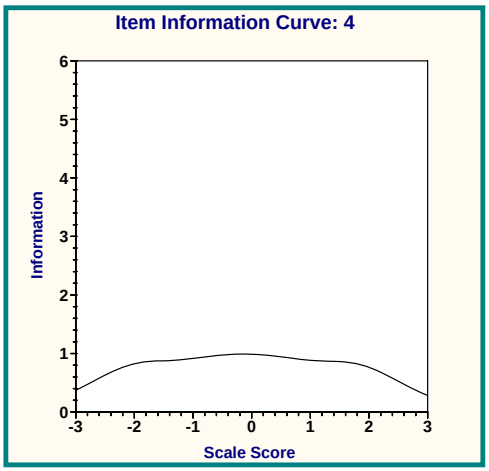
Als ik ontslag neem wordt het moeilijk om een andere baan te vinden.



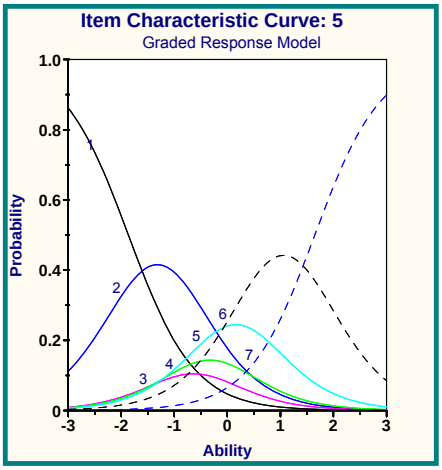


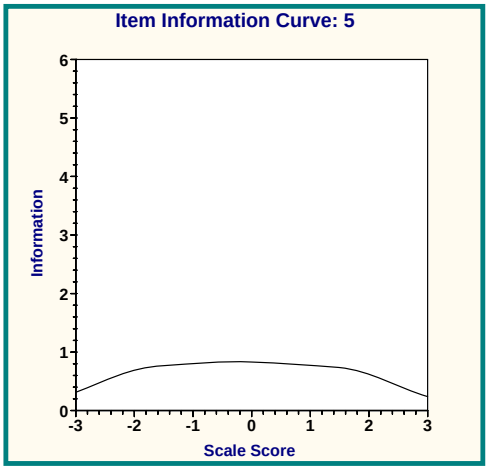
Ik heb het gevoel dat ik te weinig alternatieven heb om nu ontslag te nemen.



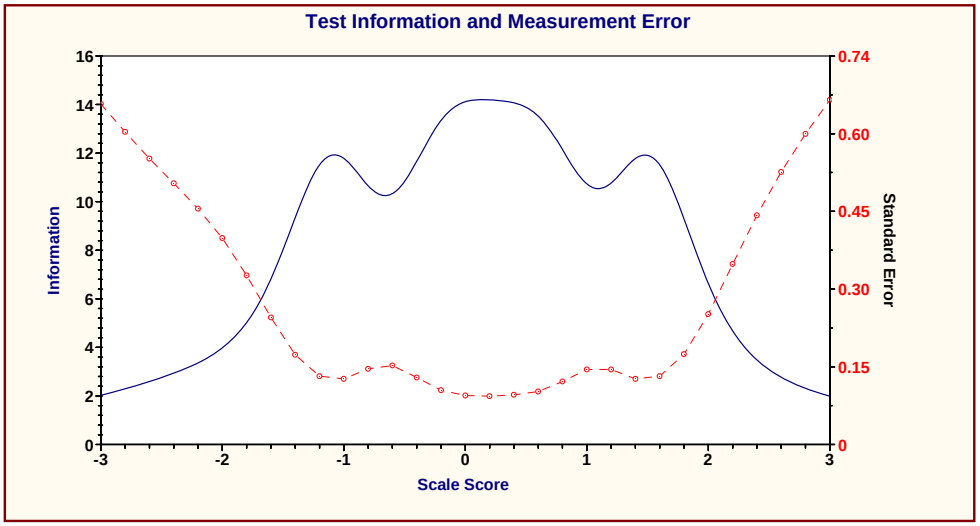


Er zou teveel in mijn leven verstoord worden als ik nu ontslag zou nemen.





Ik ben bang voor wat er zou kunnen gebeuren als ik mijn baan opzeg, zonder meteen een nieuwe baan te hebben.



Marginale betrouwbaarheid van de test: 0.9032

MULTILOG parameterschattingen van 'Continue Commitment (organisatie)'

	α (S.E.)	β1 (S.E.)	β2 (S.E.)	β3 (S.E.)	β4 (S.E.)	β5 (S.E.)	β6 (S.E.)
Item 1	1,52 (0,07)	-1,96 (0,11)	-0,50 (0,06)	-0,19 (0,05)	0,28 (0,04)	0,83 (0,07)	2,13 (0,12)
Item 2	4,14 (0,16)	-1,06 (0,04)	-0,17 (0,03)	0,04 (0,03)	0,35 (0,03)	0,65 (0,03)	1,48 (0,05)
Item 3	4,08 (0,14)	-1,16 (0,04)	-0,22 (0,03)	-0,01 (0,03)	0,34 (0,03)	0,65 (0,03)	1,55 (0,06)

Item 4	1,75 (0,08)	-1,96 (0,11)	-0,72 (0,06)	-0,40 (0,05)	-0,02 (0,05)	0,49 (0,05)	1,76 (0,10)
Item 5	1,61 (0,08)	-1,87 (0,11)	-0,77 (0,06)	-0,51 (0,06)	-0,15 (0,05)	0,47 (0,06)	1,65 (0,10)

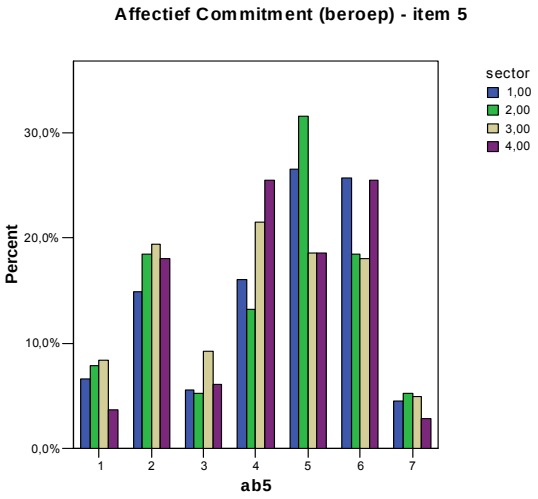
Uw houding ten aanzien van uw beroep

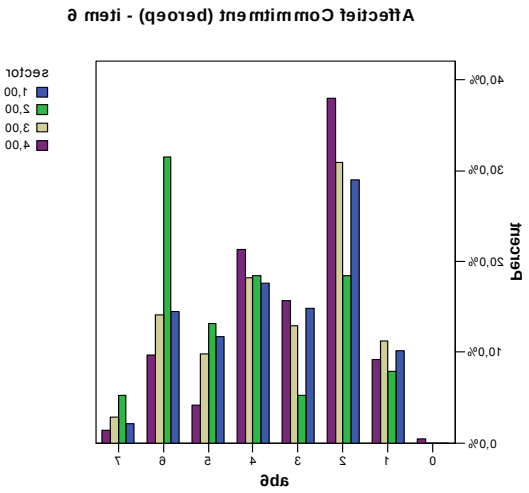
Affectief commitment

De 6 vragen uit deze schaal gaan over de mate waarin een werknemer zich verbonden voelt met zijn of haar beroep. Het gaat hier om affectief commitment naar het beroep toe.

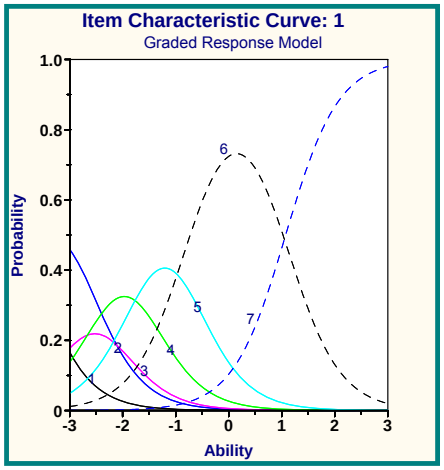
Drie van de 6 vragen moesten voor deze analyse omgevormd worden, te weten vraag 2,3 en 6. Deze vragen zijn dusdanig geformuleerd dat het antwoordpatroon omgekeerd moest worden (antwoord 1 wordt dan 7, antwoord 2 wordt 6 en zo verder).
De vragen in deze schaal zijn redelijk betrouwbaar, de betrouwbaarheid gemeten met Cronbach's alpha is 0,734. Als vraag 6 wordt weggelaten, dan wordt de alpha-waarde iets hoger, namelijk 0,753. Als item 5 en 6 weg worden weggelaten dan verkrijgen we een alpha waarde van 0,789. Als we kijken naar de IRT analyse zien we ook dat de vragen 6 (α=0,67 "Ik identificeer me niet met dit beroep") en 5 (α=0,89 "Dit beroep is belangrijk voor mijn zelfbeeld") weinig informatie verschaffen. Item 4 (α=2,71 "Ik ben enthousiast over dit beroep") geeft de meeste informatie.
Uit de testinformatie curve kunnen we aflezen dat de schaal vooral precies meet in het linker gebied (met een θ kleiner dan 0). Dit is ook goed te zien in de item informatie curves van de items 1,2,3 en 4. Deze dalen vanaf een theta van 0 sterk.
De opgetelde gemiddelde score van deze schaal is 33,33 wat een gemiddelde score inhoudt van 5,56. Dit betekent dat de meeste mensen 'mee eens' op de vragen geantwoord hebben.

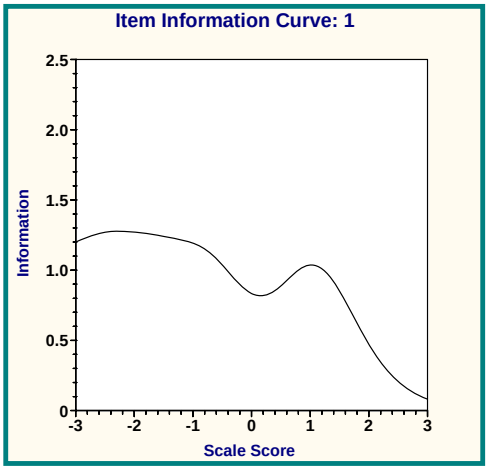
In deze schaal zijn twee vragen die zeer weinig informatie bevatten namelijk vraag 6 'Ik identificeer me niet met dit beroep' en vraag 5 'Dit beroep is belangrijk voor mijn zelfbeeld'. Vraag 5 laat alleen een klein significant verschil in invulling zien voor de respondenten van de verschillende sectoren. Vraag 6 laat ook een verschil zien in antwoorden van de respondenten in de verschillende sectoren als ook een klein verschil in antwoorden van respondenten met een verschillende opleiding. Alhoewel er op de histogrammen duidelijke verschillen te zien zijn, kan er niet één eenduidige conclusie uit de gegevens getrokken worden.



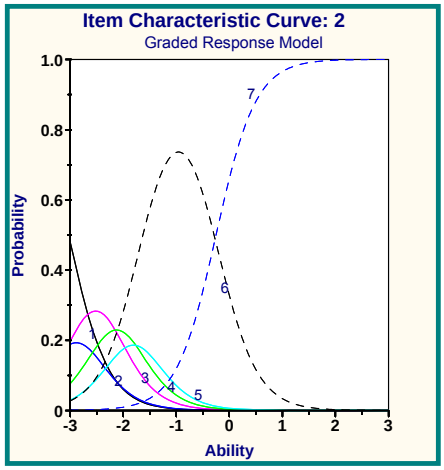


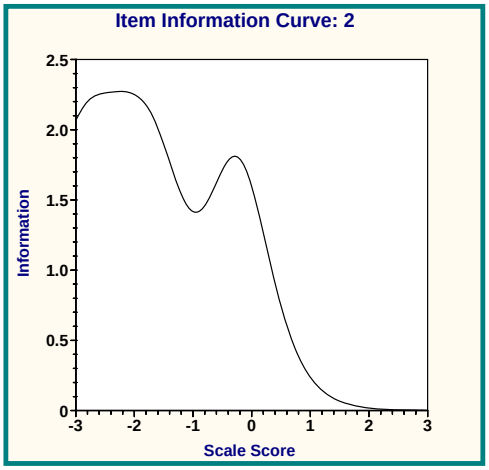
MULTILOG grafieken voor 'Affectief Commitment (beroep)'



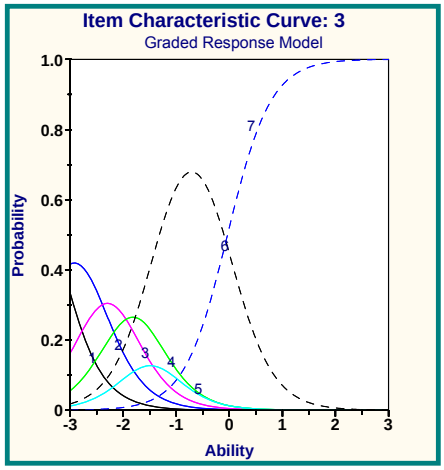


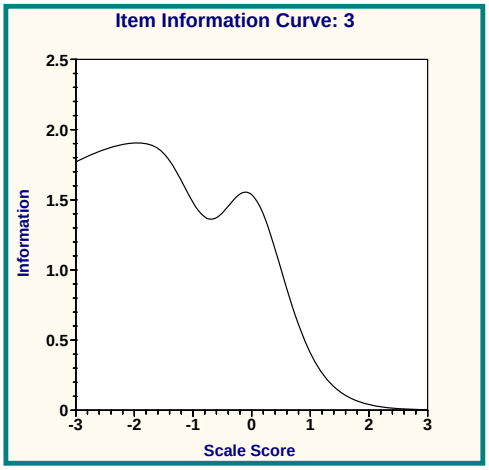
Ik ben er trots op dat ik dit beroep beoefen.



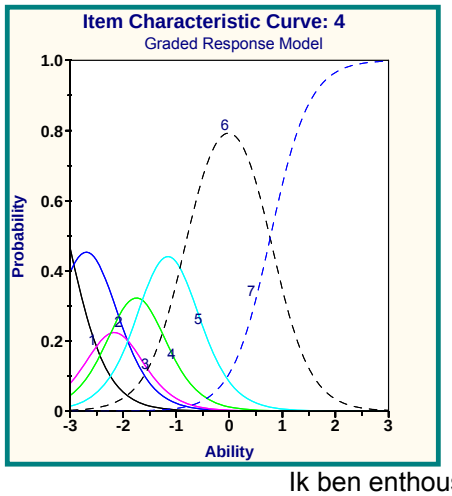


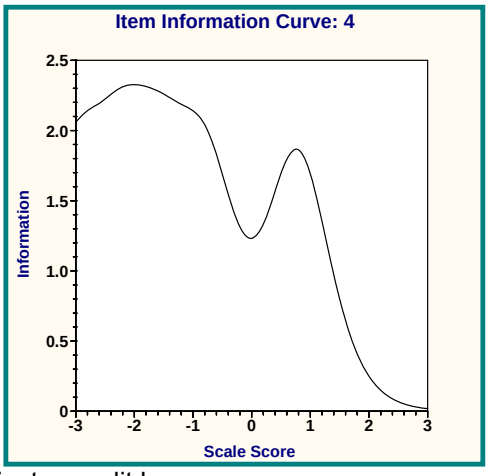
Ik heb er een hekel aan dat ik dit beroep uitoefen.



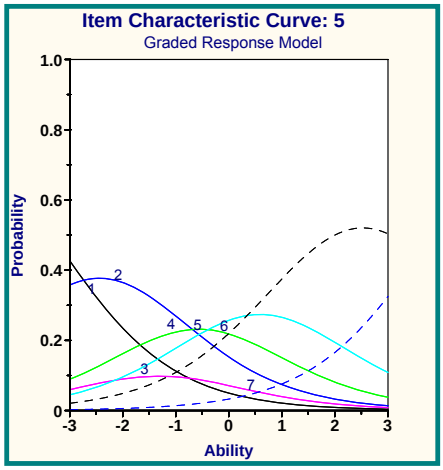


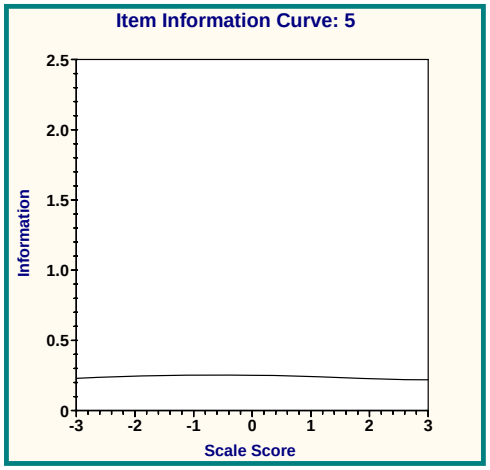
Ik heb er spijt van dat ik dit beroep gekozen heb.



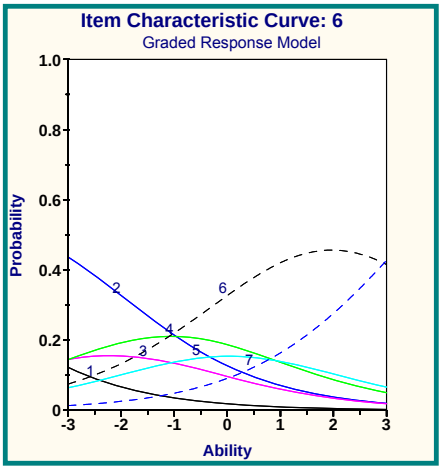


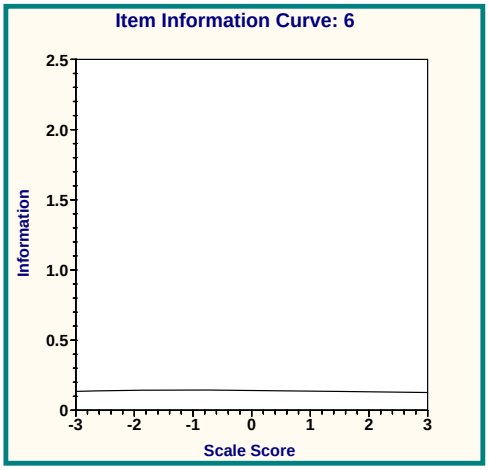
Ik ben enthousiast over dit beroep.



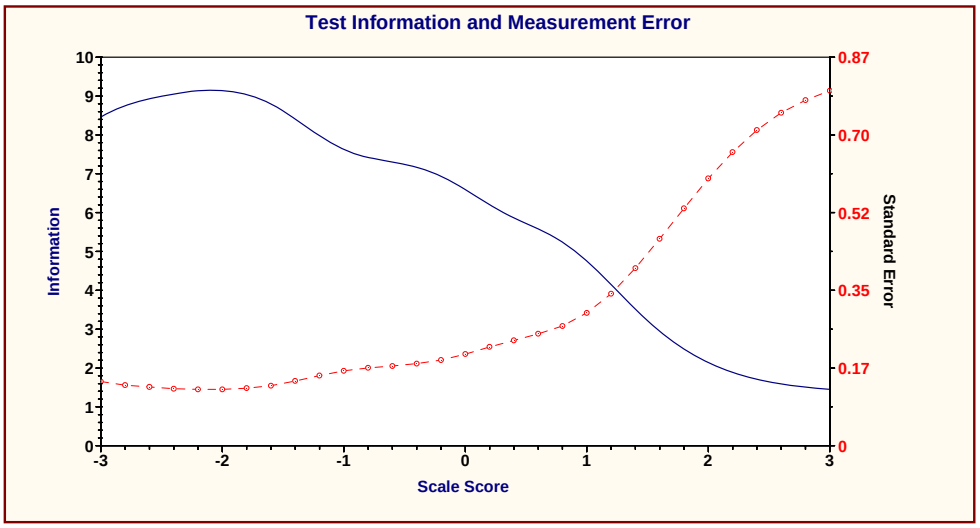


Dit beroep is belangrijk voor mijn zelfbeeld.





Ik identificeer me niet met dit beroep.



Marginale betrouwbaarheid van de test: 0,8220

MULTILOG parameterschattingen van 'Affectief Commitment (beroep)'

	α (S.E.)	β1 (S.E.)	β2 (S.E.)	β3 (S.E.)	β4 (S.E.)	β5 (S.E.)	β6 (S.E.)
Item 1	2,01 (0,10)	-3,80 (0,38)	-2,74 (0,17)	-2,30 (0,12)	-1,63 (0,08)	-0,78 (0,05)	1,08 (0,06)
Item 2	2,66 (0,16)	-3,03 (0,23)	-2,74 (0,17)	-2,30 (0,11)	-1,95 (0,09)	-1,67 (0,07)	-0,25 (0,04)
Item 3	2,45 (0,13)	-3,28 (0,29)	-2,55 (0,14)	-2,04 (0,09)	-1,60 (0,07)	-1,39 (0,06)	-0,03 (0,04)
Item 4	2,71 (0,13)	-3,06 (0,22)	-2,33 (0,11)	-2,00 (0,09)	-1,51 (0,06)	-0,81 (0,04)	0,78 (0,05)
Item 5	0,89 (0,07)	-3,33 (0,03)	-1,55 (0,15)	-1,10 (0,12)	-0,04 (0,09)	1,23 (0,12)	3,82 (0,29)
Item 6	0,67 (0,07)	-5,95 (0,70)	-2,65 (0,30)	-1,71 (0,21)	-0,43 (0,12)	0,49 (0,12)	3,44 (0,33)

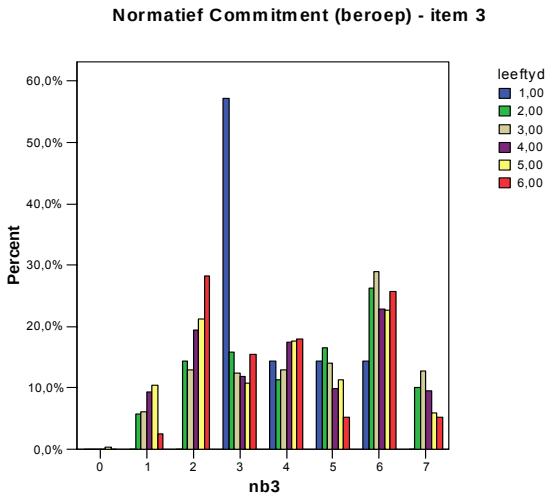
Normatief commitment

De 6 vragen uit deze schaal gaan over de mate waarin een werknemer zich verbonden voelt met zijn of haar beroep. Het gaat hier om normatief commitment naar het beroep toe.

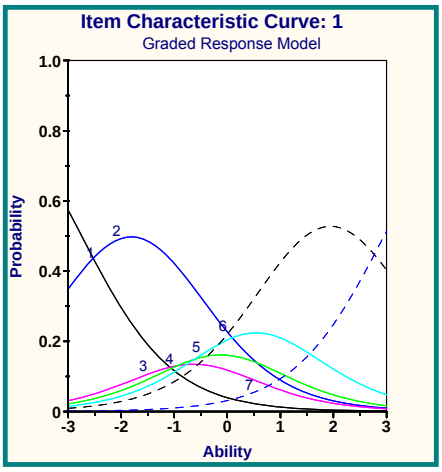
Vraag 3 van deze schaal is dusdanig gesteld ("Ik voel *geen* enkele verplichting om dit beroep te blijven uitoefenen") dat de antwoorden omgekeerd moeten worden (antwoord 1 wordt antwoord 7, antwoord 2 wordt antwoord 6 en zo verder).
De betrouwbaarheid van deze schaal is hoog, deze heeft een Cronbach's alpha van 0,806. Als we het item met de laagste item-totaal correlatie weglaten, item 3, dan wordt de alpha iets hoger, namelijk 0,814. Aan de item informatie curve van de IRT analyse is ook te zien dat item 3 (α=0,87) zeer weinig informatie geeft. Ook item 1 (α=1,16): "Ik vind dat mensen die voor een beroep zijn opgeleid, de verantwoordelijkheid hebben om dat beroep voor langere tijd uit te oefenen". Net als bij de vragen over normatief commitment naar de organisatie toe, is dit ook weer een vraag over het waarde-oordeel van anderen. De beste vragen uit deze schaal zijn vraag 4 (α=2,93):"Ik voel me verantwoordelijk ten opzichte van de beroepsgroep om dit beroep te blijven uitoefenen" en vraag 5 (α=2,46):"Ik zou me schuldig voelen als ik stopte met het uitoefenen van dit beroep".
De gemiddelde opgetelde score is 20,06 dit betekent een gemiddeld antwoord van 3,3. Een vrij lage score dus, wat inhoudt dat er veel geantwoord is in de categorieën 'niet mee eens'.
Uit de test informatie curve blijkt dat de schaal vanaf een θ van –1,5 zeer precies meet. Dit is ook weer af te lezen uit de parameterschatting tabel (voor item 1 tussen

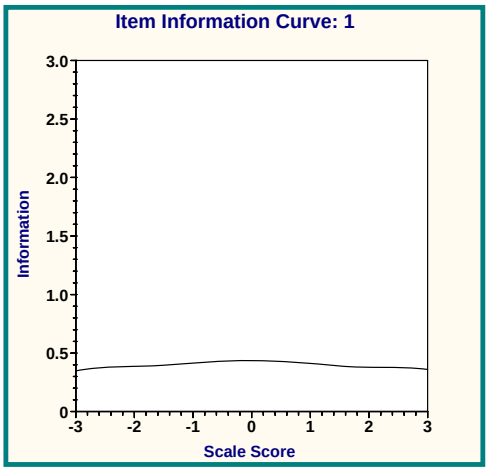
de –2,77 en 2,99; voor item 2 tussen de –0,93 en 3,33; voor item 3 tussen de –2,97 en 3,03; voor item 4 tussen de –1,18 en 2,32; voor item 5 tussen de –0,91 en 2,98 en voor item 6 tussen de –1,49 en 2,58). De schaal meet dus goed voor zowel mensen die een matig negatief normatief commitment hebben ten aanzien van hun beroep tot aan mensen die een positief normatief commitment hebben ten aanzien van hun beroep.

Item 3 ‘Ik voel geen enkele verplichting om dit beroep te blijven uitoefenen’ is de minst informatieve vraag uit deze schaal. Dit is ook goed te zien aan de informatie curve van item 3, deze loopt bijna horizontaal aan de x-as. Deze vraag toont alleen een significant verschil op de antwoorden voor de respondenten van verschillende leeftijden. Uit dit histogram is geen duidelijk onderscheid te maken, het enige wat opvalt is dat ruim 55 % van de respondenten die jonger dan 20 jaar zijn (leeftijd 1) antwoord 3 - beetje mee oneens – hebben gekozen.

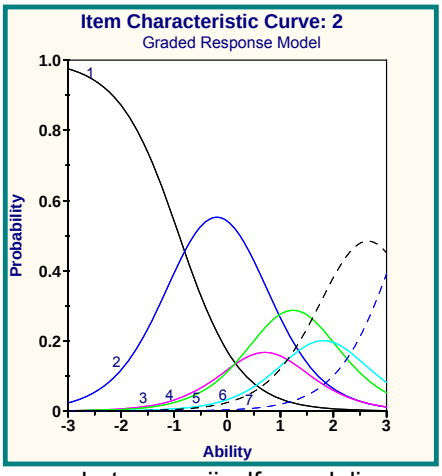


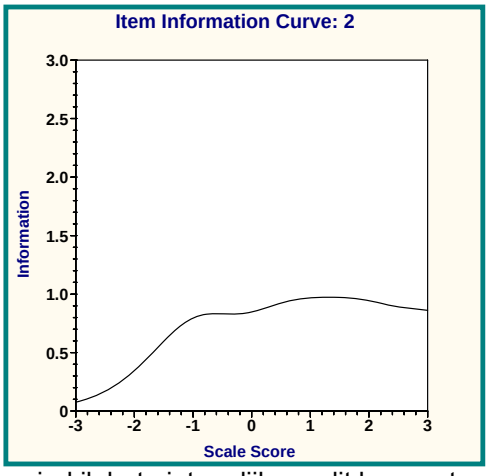
MULTILOG grafieken voor 'Normatief Commitment (beroep)'



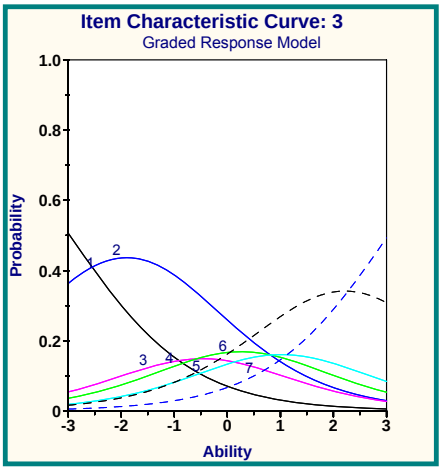


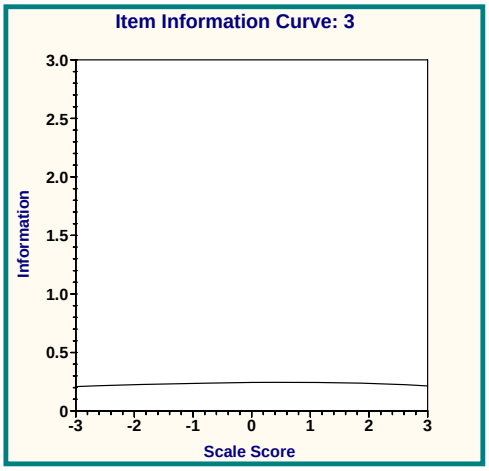
Ik vind dat mensen die voor een beroep zijn opgeleid, de verantwoordelijkheid hebben om dat beroep voor langere tijd uit te oefenen.



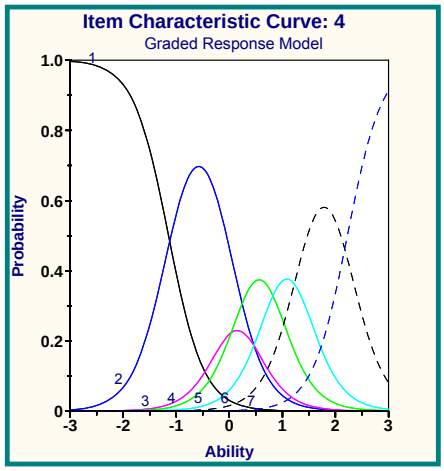


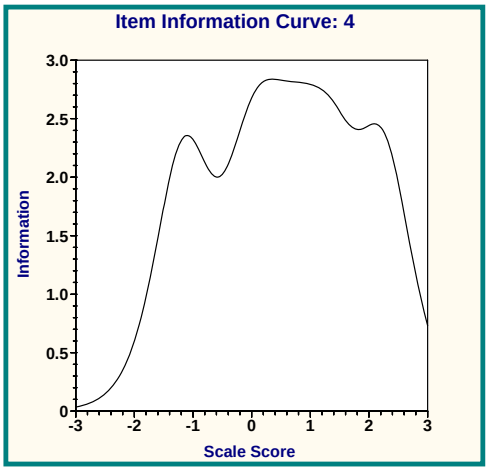
Zelfs wanneer het voor mijzelf voordelig zou zijn, vind ik het niet eerlijk om dit beroep te verlaten.



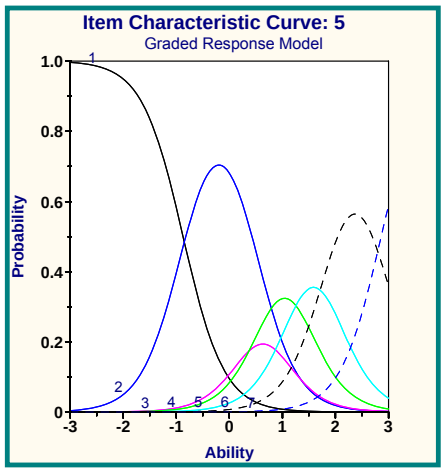


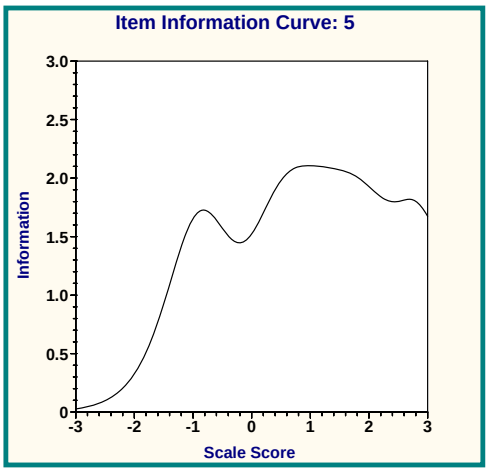
Ik voel geen enkele verplichting om dit beroep te blijven uitoefenen.



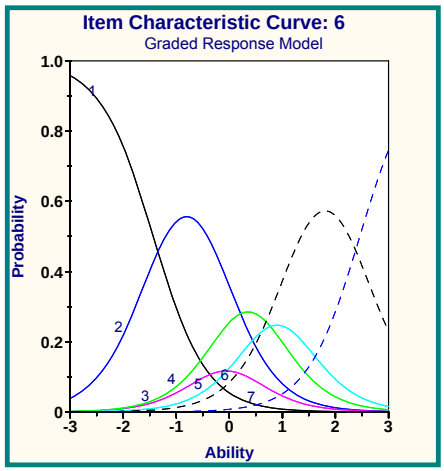


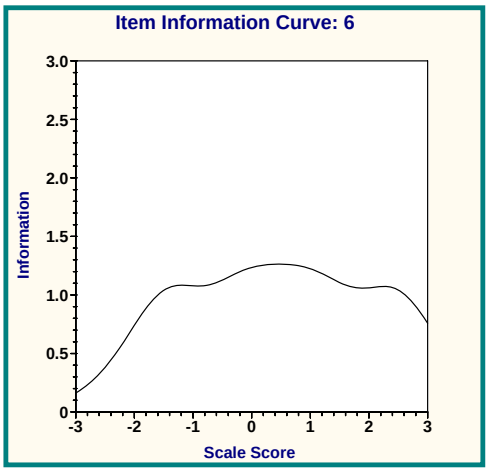
Ik voel me verantwoordelijk ten opzichte van de beroepsgroep om dit beroep te blijven uitoefenen.



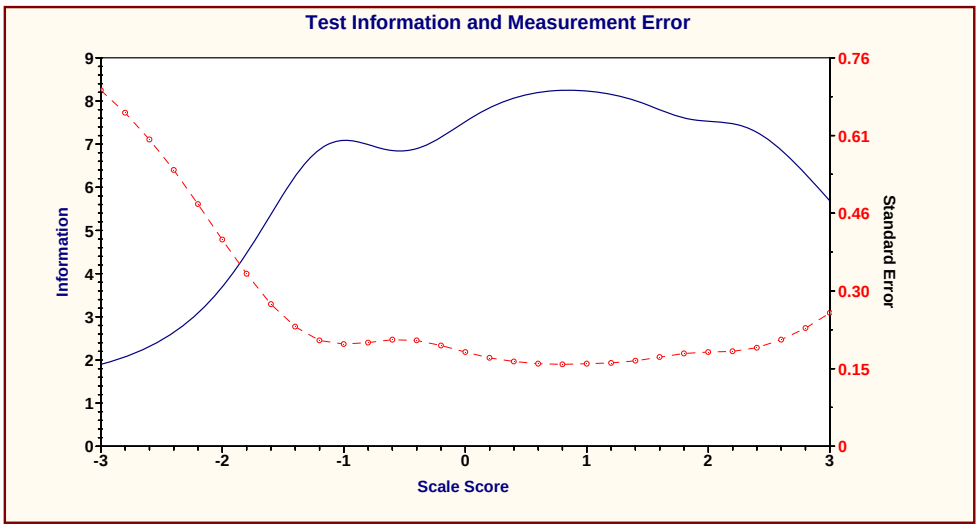


Ik zou me schuldig voelen als ik stopte met het uitoefenen van dit beroep.





Ik heb dit beroep omdat ik me loyaal voel ten opzichte van dit beroep.



Marginale betrouwbaarheid van de test: 0,8578.

MULTILOG parameterschattingen van 'Normatief Commitment (beroep')

	α (S.E.)	β1 (S.E.)	β2 (S.E.)	β3 (S.E.)	β4 (S.E.)	β5 (S.E.)	β6 (S.E.)
Item 1	1,16 (0,08)	-2,77 (0,18)	-0,88 (0,08)	-0,41 (0,07)	0,16 (0,07)	0,95 (0,09)	2,99 (0,22)
Item 2	1,73 (0,08)	-0,93 (0,07)	0,52 (0,05)	0,92 (0,06)	1,61 (0,09)	2,09 (0,12)	3,33 (0,24)
Item 3	0,87 (0,06)	-2,97 (0,23)	-0,82 (0,11)	-0,13 (0,09)	0,65 (0,10)	1,40 (0,14)	3,03 (0,28)
Item 4	2,93 (0,11)	-1,18 (0,05)	-0,01 (0,03)	0,31 (0,04)	0,85 (0,04)	1,40 (0,06)	2,31 (0,12)
Item 5	2,46 (0,11)	-0,91 (0,05)	0,50 (0,04)	0,81 (0,05)	1,35 (0,06)	1,96 (0,09)	2,98 (0,23)
Item 6	1,89 (0,08)	-1,49 (0,08)	-0,19 (0,05)	0,06 (0,05)	0,67 (0,05)	1,20 (0,07)	2,58 (0,16)

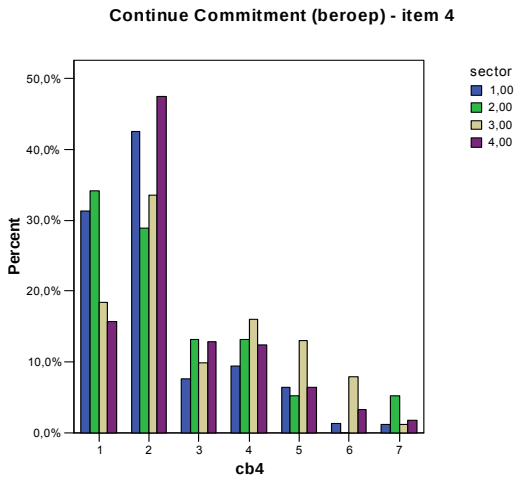
Continue commitment

Ook deze 6 vragen gaan over de mate waarin een werknemer zich verbonden voelt met zijn of haar beroep. Het gaat hier echter om continue commitment naar het beroep toe.

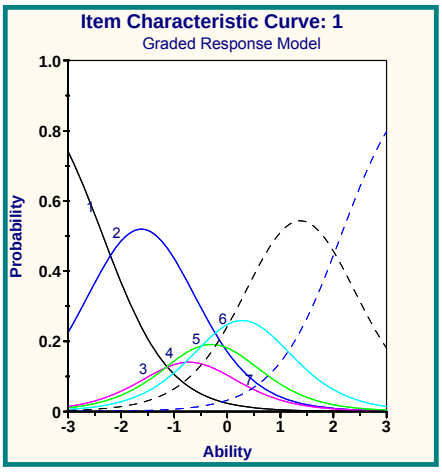
De betrouwbaarheid van deze test is gemeten met Cronbach's alpha 0,822. Item 4 heeft de kleinste item-totaal correlatie en als dit item wordt weggelaten stijgt de waarde van alpha nauwelijks (wordt 0,824). In de IRT analyse zie je dit ook goed aan de item informatie curve van item 4 (α= 1,12 "Ik voel me gedwongen om dit beroep te

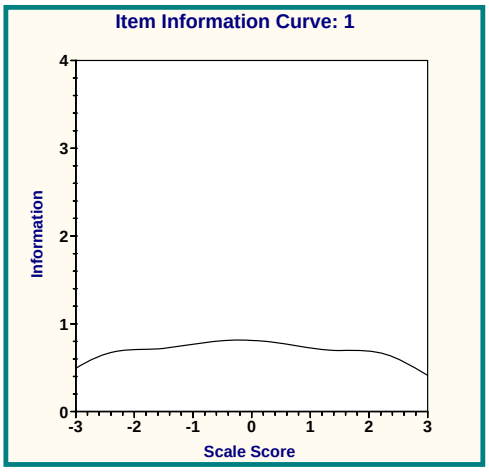
blijven uitoefenen”). Ook de items 1 met een α van 1,59 (“Van beroep veranderen zou op dit moment te moeilijk voor mij zijn”) en 2 met een α van 1,17 (“Ik heb teveel geïnvesteerd in dit beroep om te overwegen van beroep te veranderen”) leveren weinig informatie. Het item dat de meeste informatie geeft is item 6 (α=3,16 “Het zou me teveel kosten als ik nu van beroep zou veranderen”).
Het gemiddelde opgetelde antwoord van deze schaal is 21,68 wat neerkomt op een gemiddelde antwoord van 3,6. Dit neigt wat meer naar de “niet mee eens” antwoorden.
De test informatie curve laat zien dat de schaal precies meet op het gebied van een θ die tussen de –2 en +2 ligt. Alleen voor heel lage waarden van θ (mensen met een heel negatieve houding naar hun beroep toe) en voor heel hoge waarden van θ (mensen met een heel positieve houding naar hun beroep toe) meet de schaal minder nauwkeurig en neemt de meetfout toe.

Item 4 ‘Ik voel me gedwongen om dit beroep te blijven uitoefenen’ is het minst informatieve item van deze schaal. De verschillen in antwoordpatroon van de respondenten is alleen voor mannen en vrouwen gelijk. Voor de verschillende opleidingen, leeftijden, werktijd in jaren en sector waar de respondenten werkzaam zijn is er een significant verschil waar te nemen. Het verschil tussen de verschillende sectoren is het grootste, deze wordt dan ook toegelicht middels onderstaand histogram. Opvallend is dat vooral mensen die werkzaam zijn in het ziekenhuis en in de zorginstelling zich niet gedwongen voelen, terwijl respondenten die werkzaam zijn in de petro chemische industrie en de vrijwilligers van de zorginstelling zich vaker gedwongen voelen om het beroep uit te blijven oefenen.

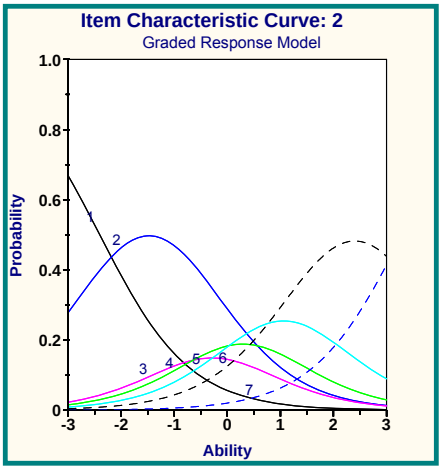


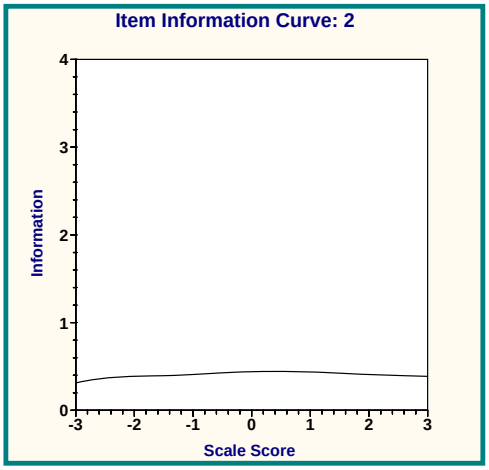
MULTILOG grafieken voor 'Continue Commitment (beroep)'



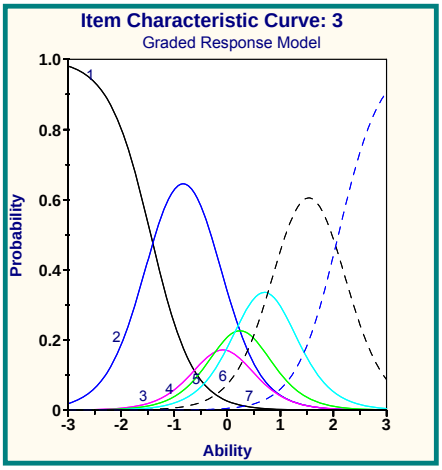


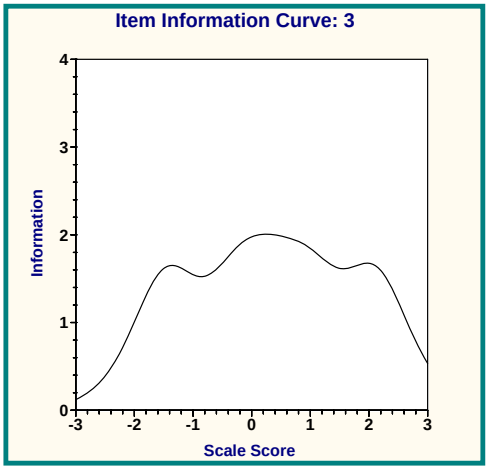
Van beroep veranderen zou op dit moment te moeilijk voor me zijn.



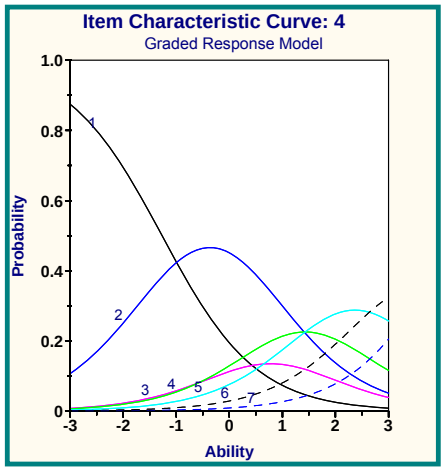


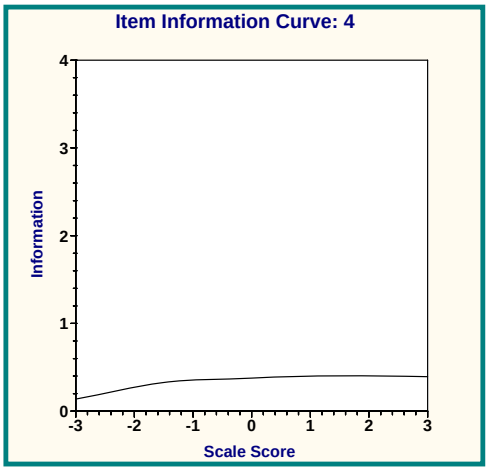
Ik heb teveel geïnvesteerd in dit beroep om te overwegen van beroep te veranderen.



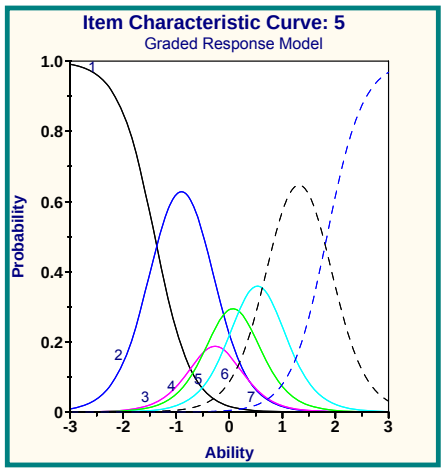


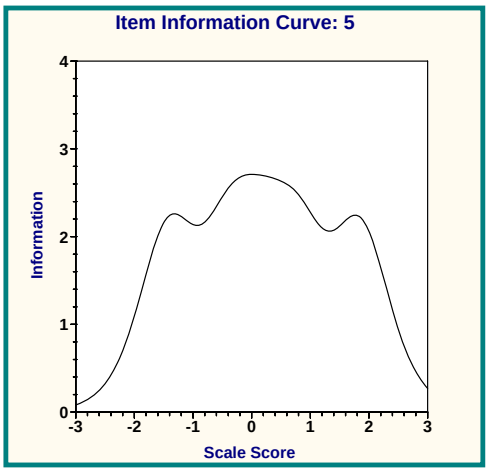
Er zou teveel in mijn leven verstoord worden als ik van beroep zou veranderen.



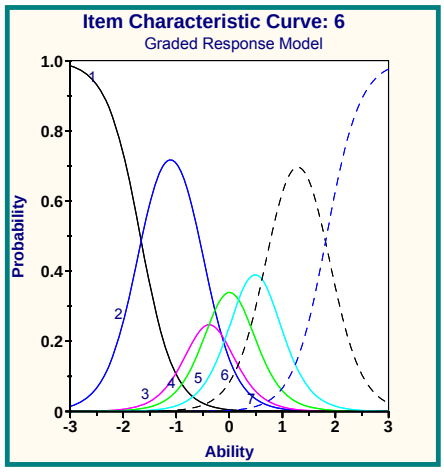


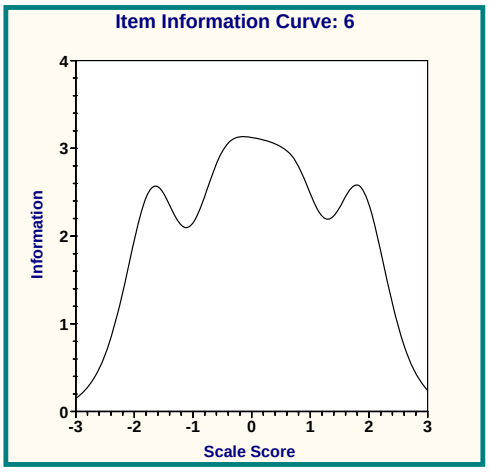
Ik voel me gedwongen om dit beroep te blijven uitoefenen.



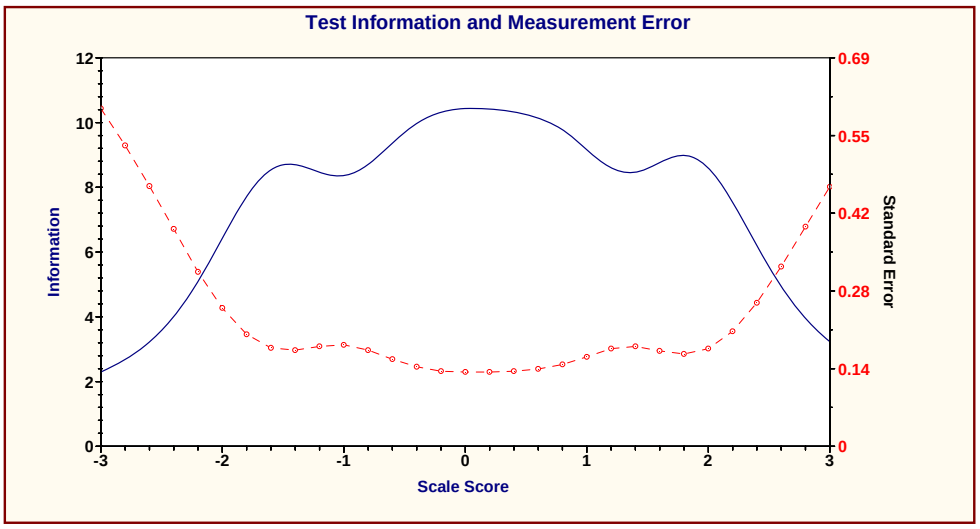


Het zou op dit moment forse persoonlijke offers van me vergen wanneer ik een ander beroep zou kiezen.





Het zou me teveel kosten als ik nu van beroep zou veranderen.



Marginale betrouwbaarheid van de test: 0.8900

MULTILOG parameterschattingen voor 'Continue Commitment (beroep)'

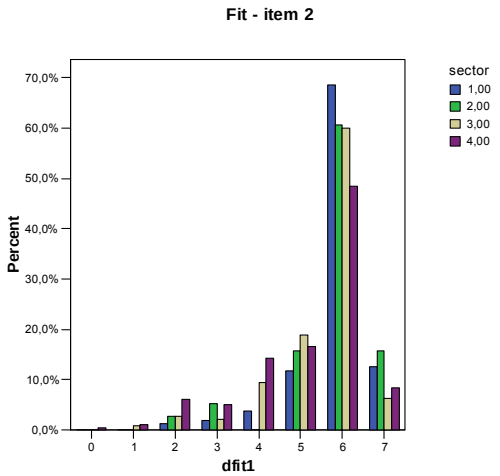
	α (S.E.)	β1 (S.E.)	β2 (S.E.)	β3 (S.E.)	β4 (S.E.)	β5 (S.E.)	β6 (S.E.)
Item 1	1,59 (0,07)	-2,34 (0,14)	-0,90 (0,07)	-0,54 (0,06)	-0,06 (0,05)	0,61 (0,06)	2,14 (0,13)
Item 2	1,17 (0,07)	-2,41 (0,16)	-0,55 (0,08)	-0,03 (0,07)	0,62 (0,08)	1,50 (0,11)	3,30 (0,24)
Item 3	2,51 (0,10)	-1,44 (0,07)	-0,22 (0,04)	0,05 (0,04)	0,42 (0,04)	0,98 (0,05)	2,09 (0,10)
Item 4	1,12 (0,07)	-1,27 (0,12)	0,53 (0,07)	1,02 (0,09)	1,84 (0,13)	2,90 (0,20)	4,21 (0,35)
Item 5	2,93 (0,11)	-1,41 (0,06)	-0,40 (0,04)	-0,14 (0,04)	0,27 (0,04)	0,79 (0,04)	1,84 (0,08)
Item 6	3,16 (0,12)	-1,68 (0,07)	-0,54 (0,04)	-0,22 (0,03)	0,23 (0,03)	0,75 (0,04)	1,83 (0,08)

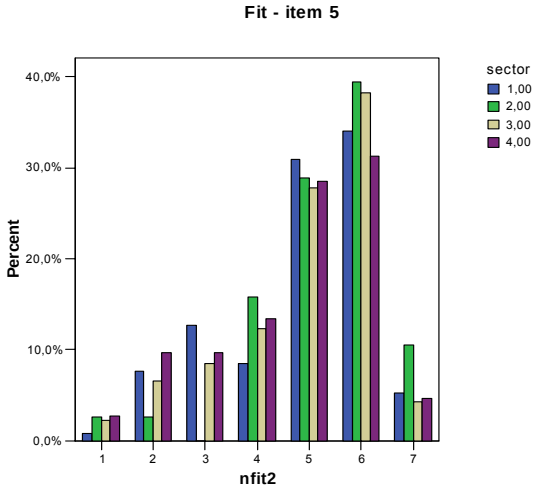
Hoe u vindt dat uw organisatie en u bij elkaar passen

De 9 vragen uit deze categorie zijn bedoeld om een beeld te krijgen van de mening van de respondenten over hoe goed ze de organisatie en zichzelf bij elkaar vinden passen. De vragen zijn onderverdeeld in 3 categorieën: persoon-organisatie fit (items 1,5 en 9), persoon-baan fit (items 2,3 en 8) en behoefte voorziening (items 4,6 en 7). Ze worden echter als 1 schaal behandeld.

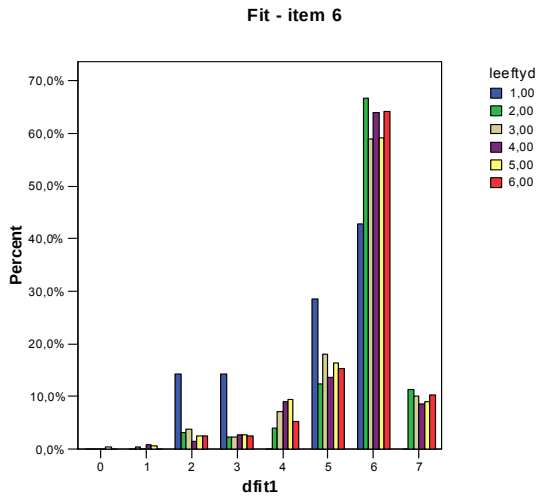
De betrouwbaarheid van deze schaal is hoog, de waarde van Cronbach's alpha is 0,892. Er zitten geen slechte items tussen, de betrouwbaarheid kan dan ook niet vergroot worden door een van de items weg te laten.
Als we naar de IRT analyse kijken valt op dat de items 1 (α=1,37), 5 (α=1,34) en 9 (α-1,36) de minste informatie opleveren. Dit zijn de vragen over hoe de respondenten vinden dat ze bij de organisatie passen. De vragen over hoe de respondenten vinden dat ze bij hun beroep passen leveren iets meer informatie op (alpha waarden van 1,86, 2,17 en 2,04). De vragen over de verwachtingen van de baan (behoefte) leveren de meeste informatie met alpha waarden van 3,42 voor item 4; 3,11 voor item 6 en 3,60 voor item 7. Uit de gegevens is goed te zien dat er eigenlijk 3 keer dezelfde vraag in een andere formulering gesteld is.
Uit de testinformatie curve maken we op dat de test voor een θ waarde onder de 0 precies meet, bij een θ waarde van 1 daalt de precisie en stijgt de standaardmeetfout (dit omdat mensen met een θ van 1 een zeer hoge kans hebben om antwoord 6 (mee eens) te geven). Vanaf een θ waarde van 2 loopt de informatie curve weer wat op, om daarna bij een nog hogere θ waarde weer omlaag te gaan. Uit de parameterschatting tabel kun je ook aflezen dat de schaal over een groot deel van het theta continuüm meet. Voor de vragen 1,5 en 9 van ongeveer –3,5 tot +3,5; voor de vragen 2,3 en 8 van ongeveer –3 tot +1,5 en voor de vragen 4,6 en 7 van ongeveer –2,5 tot 2.

Per onderdeel van deze schaal kijken we nog eens naar het minst informatieve item, dit is voor persoon-baan fit item 2, voor persoon-organisatie fit item 5, en voor behoefte voorziening item 6. Voor item 2 (gezien de vaardigheden die voor mijn werk vereist zijn, pas ik daar heel goed bij) is er alleen een significant verschil in antwoordenpatroon tussen de mensen van de verschillende sectoren. Het opvallendste verschil hierin is dat respondenten die in het ziekenhuis werken het minder vaak met deze stelling eens zijn dan de respondenten die in de zorginstelling of in de petro chemische industrie werken. Ook voor item 5 (de zaken waar ik veel waarde aan hecht in het leven komen sterk overeen met de zaken waar mijn organisatie veel waarde aan hecht) geldt dat er alleen een significant verschil aanwezig is in antwoordpatroon voor respondenten uit de verschillende sectoren.

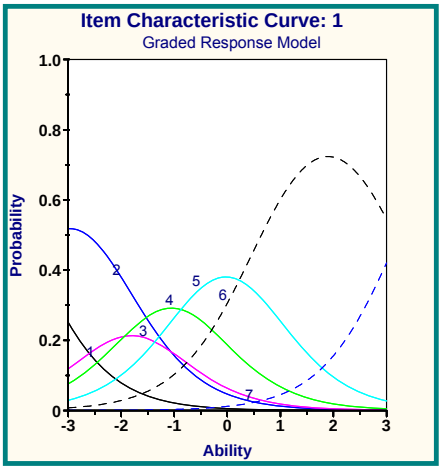


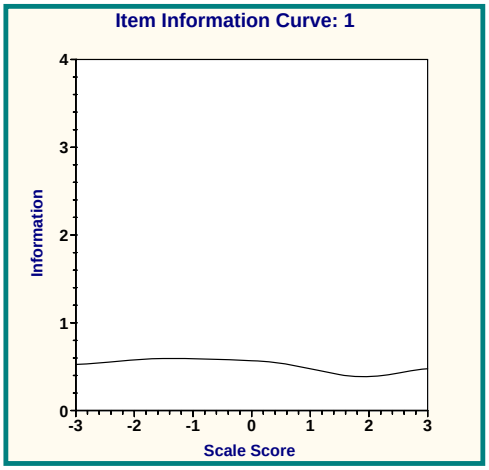


Ten slotte item 6 (mijn huidige baan sluit precies aan bij de eisen die ik aan een baan stel). Het enige significante verschil wat bij deze vraag naar voren komt is dat tussen respondenten van verschillende leefijden. Het opvallendste verschil hierin is dat de jonge mensen (jonger dan 20 (1)) het vaker met deze stelling oneens zijn dan de mensen die ouder dan 20 jaar zijn.

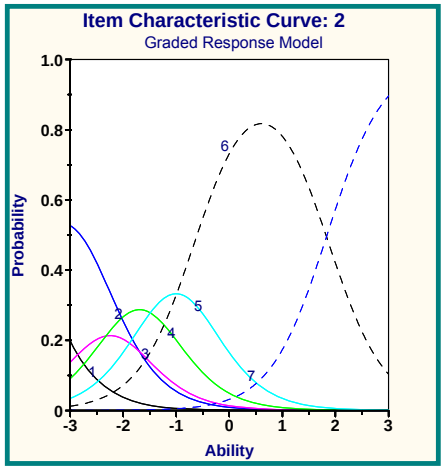


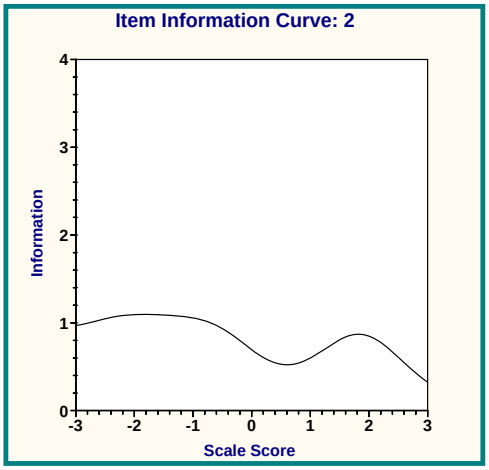
MULTILOG grafieken voor 'hoe u en uw organisatie bij elkaar passen'



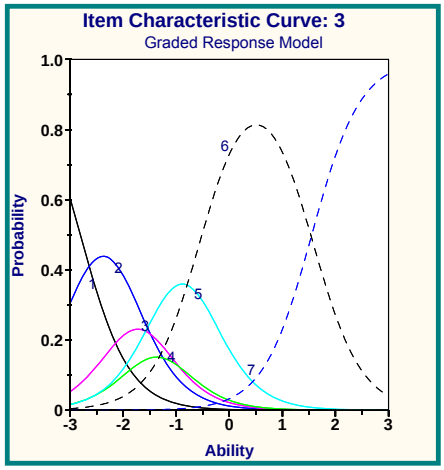


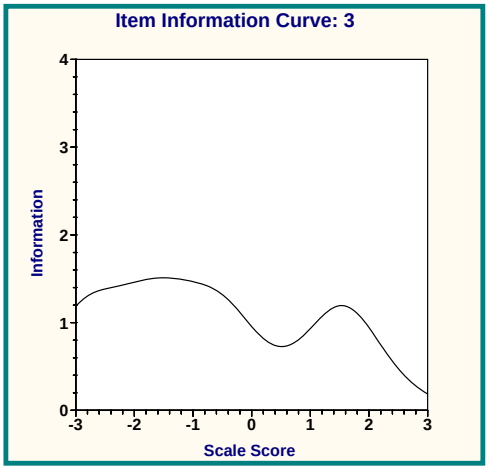
Mijn persoonlijke waarden komen overeen met de waarden en cultuur van mijn organisatie.



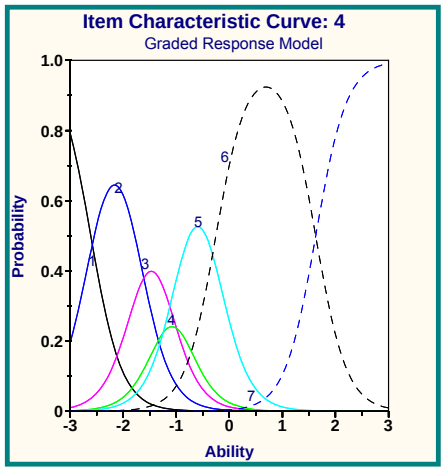


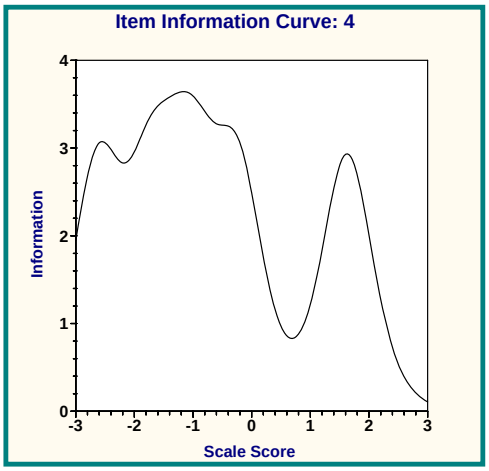
Gezien de vaardigheden die voor mijn werk vereist zijn, pas ik daar heel goed bij.



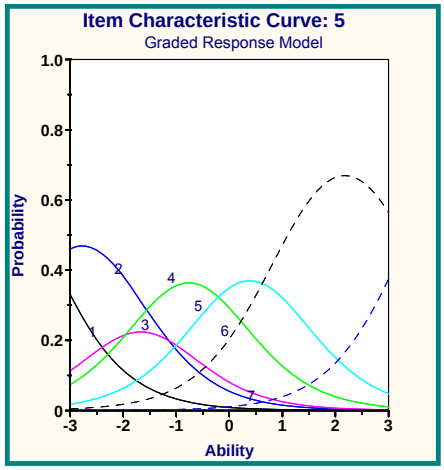


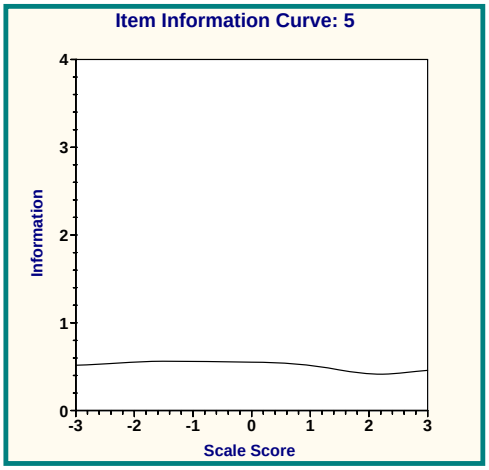
Mijn talenten en scholing passen goed bij de eisen die mijn baan aan mij stelt.



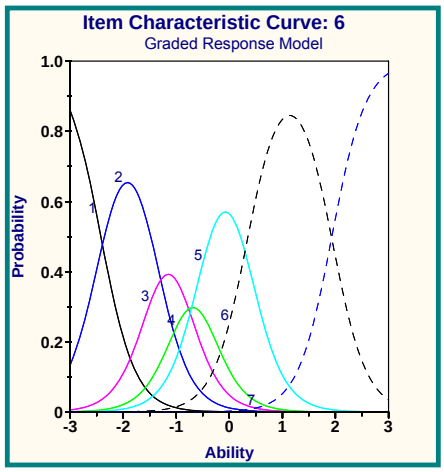


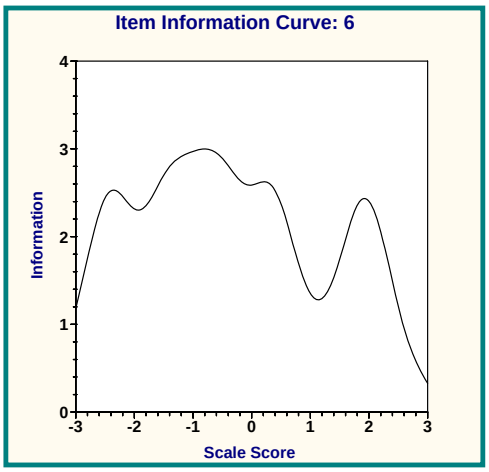
Wat mijn baan me biedt, past goed bij wat ik van een baan verwacht.



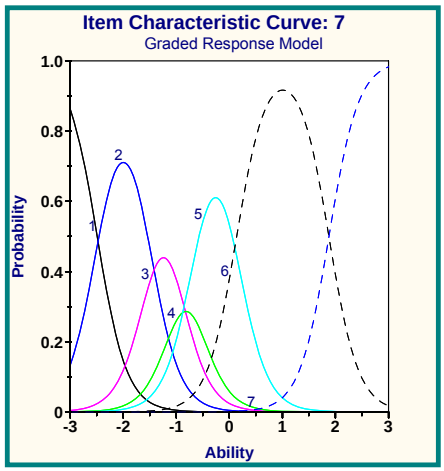


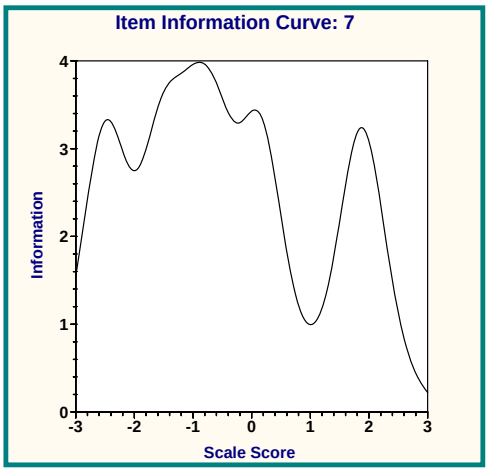
De zaken waar ik veel waarde aan hecht in het leven komen sterk overeen met de zaken waar mijn organisatie veel waarde aan hecht.



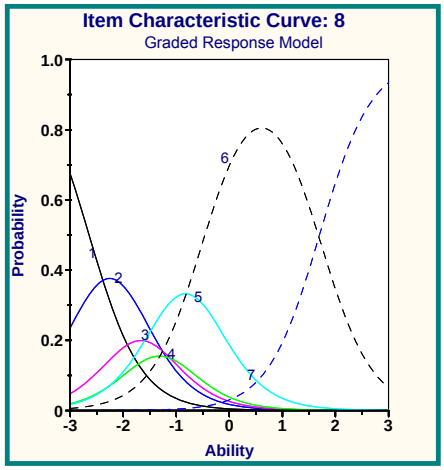


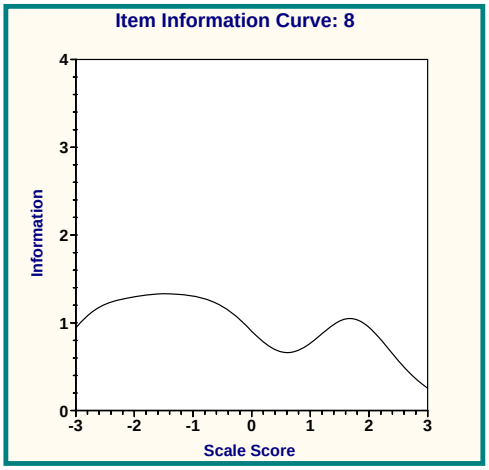
Mijn huidige baan sluit precies aan bij de eisen die ik aan een baan stel.



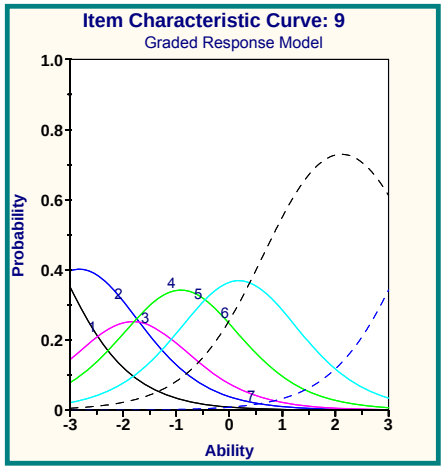


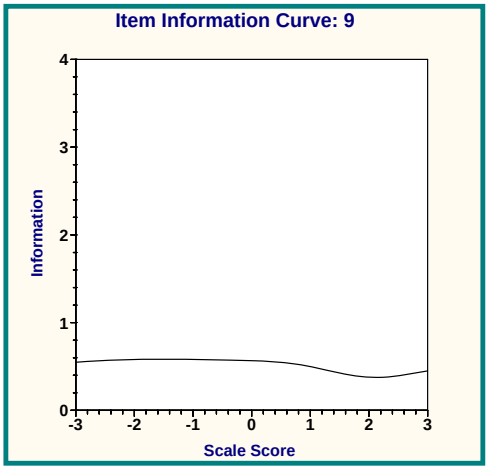
De kenmerken die een baan volgens mij moet hebben, worden heel goed vervuld in mijn huidige baan.



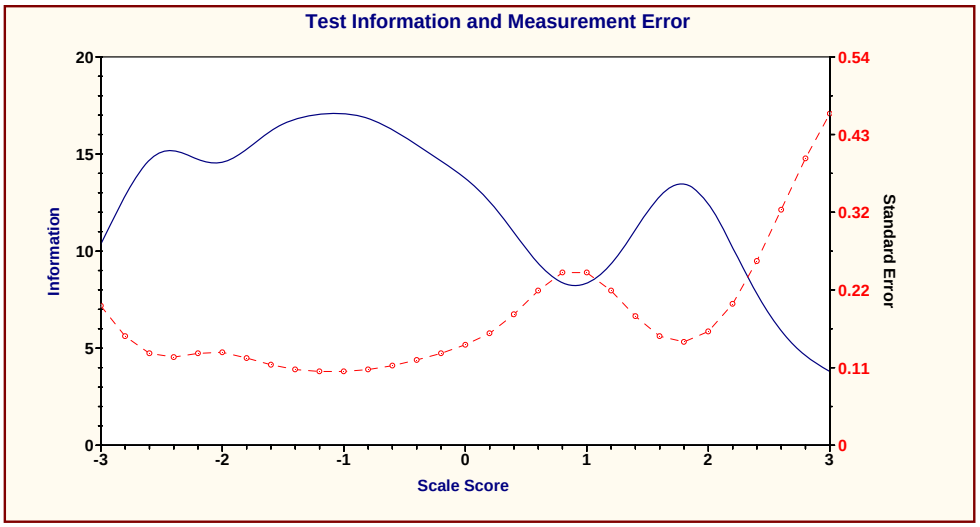


Mijn vaardigheden en opleiding passen goed bij de eisen van mijn baan.





De waarden en cultuur van mijn organisatie sluiten goed aan bij wat ik waardevol en belangrijk vind in het leven.



Marginale betrouwbaarheid van de test: 0,9194.

MULTILOG parameterschattingen voor 'hoe u en uw organisatie bij elkaar passen'

	α (S.E.)	β1 (S.E.)	β2 (S.E.)	β3 (S.E.)	β4 (S.E.)	β5 (S.E.)	β6 (S.E.)
Item 1	1,37 (0,08)	-3,80 (0,31)	-2,12 (0,14)	-1,49 (0,10)	-0,61 (0,07)	0,56 (0,06)	3,23 (0,21)
Item 2	1,86 (0,10)	-3,75 (0,34)	-2,48 (0,16)	-2,01 (0,11)	-1,38 (0,08)	-0,63 (0,05)	1,84 (0,10)
Item 3	2,17 (0,11)	-2,81 (0,18)	-1,94 (0,10)	-1,51 (0,08)	-1,23 (0,07)	-0,54 (0,05)	1,55 (0,08)
Item 4	3,42 (0,16)	-2,61 (0,13)	-1,72 (0,07)	-1,23 (0,05)	-0,94 (0,04)	-0,25 (0,03)	1,63 (0,06)
Item 5	1,34 (0,08)	-3,53 (0,27)	-2,01 (0,13)	-1,33 (0,10)	-0,19 (0,06)	0,97 (0,08)	3,38 (0,23)
Item 6	3,11 (0,13)	-2,42 (0,12)	-1,42 (0,06)	-0,88 (0,04)	-0,49 (0,04)	0,34 (0,03)	1,93 (0,08)
Item 7	3,60 (0,16)	-2,49 (0,12)	-1,50 (0,06)	-0,98 (0,04)	-0,65 (0,04)	0,14 (0,03)	1,88 (0,07)
Item 8	2,04 (0,10)	-2,65 (0,16)	-1,87 (0,10)	-1,47 (0,08)	-1,16 (0,07)	-0,48 (0,05)	1,70 (0,09)
Item 9	1,36 (0,08)	-3,45 (0,26)	-2,20 (0,15)	-1,44 (0,10)	-0,39 (0,07)	0,75 (0,07)	3,48 (0,24)

Discussie en Conclusie

Het doel van deze studie was om met behulp van Item Respons Theorie te evalueren of de verschillende schalen van de medewerkerstevredenheids vragenlijst en de individuele vragen uit de categorieën 'vaardigheden van de organisatie', 'uw vaardigheden', 'uw houding ten op zichtte van de organisatie (affectief, normatief, continue)', 'uw houding ten aanzien van uw beroep (affectief, normatief, continue)' en 'hoe vindt u dat u en de organisatie bij elkaar passen' bruikbaar zijn.
De literatuur over 'vaardigheden van de organisatie', 'commitment' en 'fit' is behandeld, evenals een inleiding in de Item Respons Theorie en een beschrijving van de verschillende typen dichotome en polytome IRT modellen. Ook het schatten van item- en vaardigheids parameters bij dichotome modellen is behandeld.
Samejima's (1969) Graded Respons Model is gebruikt om de gegevens te analyseren. Onderzoek van de item karakteristieke curves laat zien dat bijna alle vragen voldoende informatie geven, maar dat er in elke schaal ook wel 1 of 2 vragen tussen zitten die minder informatief zijn. Zeker met het oog op de relatief lange vragenlijst, zou het beter zijn om van elke schaal de 1 of 2 minst informatieve vragen te verwijderen. Om te kijken of de minst informatieve vragen echt weinig bruikbaar zijn, is er voor deze vragen ook nog een variantie analyse uitgevoerd. De 9 verschillende schalen worden hieronder één voor één besproken.

De schaal 'vaardigheden van de organisatie' is een goed bruikbare schaal, met een hoge betrouwbaarheid. Alle vragen geven voldoende informatie en meten over een breed deel van het theta-continuüm.
Voor deze schaal is het minst informatie item item 2 (Ik vind dat de organisatie in staat is om steeds nieuwe ideeën te ontwikkelen). Omdat er hier alleen een significant verschil in antwoordpatroon is geconstateerd tussen de respondenten van de verschillende sectoren (personeel uit een zorginstelling (vast en vrijwillig), petro-chemische fabrieks medewerkers en werknemers uit het ziekenhuis) kan opgemaakt worden dat deze vraag echt zeer weinig informatie verschaft, immers bij een volgend onderzoek zijn de sectoren anders en kan daar weinig algemene informatie uitgehaald worden. Deze vraag zou dan ook uit de schaal verwijderd kunnen worden, zodat de schaal iets verkort wordt.

Over de schaal 'uw vaardigheden' kunnen we zeggen dat deze schaal betrouwbaar is, maar niet nauwkeurig meet bij de hogere latente vaardigheden. En juist daar zou hij gezien het antwoordpatroon met een gemiddelde van bijna 6 juist goed moeten meten. Verder zien we dat hoe hoger de latente vaardigheid, hoe zekerder men is van zijn eigen kunnen.
Voor deze schaal is item 7 (ik ben in staat om heel ordelijk te werken) het minst informatieve item. Door de variantie analyse zien we dat er wel verschil is tussen de respondenten, met het grootste verschil tussen de respondenten met verschillende opleiding. Mensen met een hogere opleiding antwoorden vaker positief op de vraag. Maar ook in deze vraag is goed te zien dat de tendens van deze schaal bij de 'mee

eens' antwoorden ligt. Deze vraag zou dus -met het oog op de lengte van de vragenlijst- zonder verlies van resultaten weggelaten kunnen worden.

De schaal 'affectief commitment naar de organisatie toe' geeft een hoge betrouwbaarheid, zeker als item 5 wordt weggelaten. Hij meet goed in het middelste gebied van het theta continuüm, dat wil zeggen bij respondenten die een gemiddelde houding hebben ten opzichte van hun organisatie.
Item 5 laat echter wel iets opvallends zien: het blijkt dat de respondenten uit sector 3 (petro chemische industrie) veel vaker aangeven problemen van de organisatie als hun eigen problemen te zien dan mensen die in een ziekenhuis of zorginstelling werken. Zo ook geldt dit voor oudere respondenten, die zien problemen van hun organisatie ook veel meer als hun eigen problemen dan de jongere respondenten. Uit de IRT analyse blijkt dat deze vraag echt heel weinig informatie geeft (te zien aan de bijna horizontale lijn vlak langs de x-as), maar uit verder onderzoek levert deze vraag toch wel interessante gegevens op.

De betrouwbaarheid van de schaal 'normatief commitment naar de organisatie toe' is met 0,727 aan de lage kant, de schaal meet echter wel op het hele gebied van de theta: dus van zowel respondenten met een negatieve danwel een positieve houding naar de organisatie toe.
De vraag 'ik vind dat iemand loyaal zou moeten zijn ten opzichte van zijn of haar organisatie' (nummer 5) geeft de minste informatie in deze schaal volgens de IRT analyse. Maar deze vraag geeft wel een inzicht in de verschillende antwoorden per leeftijdscategorie. Zo blijkt bijvoorbeeld dat oudere werknemers meer waarde hechten aan loyaliteit dan jongere werknemers.
Deze schaal is de slechtst passende schaal in dit onderzoek: de verplichting naar de organisatie toe om te blijven. Deze schaal bevat een drietal items die een waarde-oordeel over andere mensen vragen. Deze items zijn weinig informatief, blijkbaar vinden mensen dat ze niet kunnen oordelen over of voor iemand anders. Bij de schaal normatief commitment (beroep) zit ook een dergelijk item dat een waarde oordeel over anderen vraagt dat weinig informatie geeft. Dit soort vragen zou dus beter weggelaten kunnen worden in een vragenlijst.
Hoewel de betrouwbaarheid van deze schaal dus niet erg hoog is en het blijkt dat vragen over anderen weinig informatief zijn, kan nader onderzoek van de vragen wel meer inzicht geven in het verschil in normatief commitment naar de organisatie toe in de verschillende leeftijdsgroepen.

De schaal 'continue commitment naar de organisatie toe' meet goed, daar waar hij goed meten moet, namelijk voor een θ tussen de –1,5 en de 1,5 (voor mensen met een gematigde negatieve houding ten opzichte van de organisatie tot mensen met een gematigde positieve houding). Dit valt ook af te lezen aan de hoge betrouwbaarheid van 0,835. Voor mensen met een zeer negatieve houding dan wel een zeer positieve houding is deze schaal minder nauwkeurig in zijn meten.
Uit de variantie analyse blijkt duidelijk dat het continue commitment naar de organisatie toe sterk afhangt van leeftijd en de tijd dat men voor een organisatie werkt. Dit geldt voor alle vragen in deze schaal en daarom kan item 1 'Het zou voor

mij op dit moment moeilijk zijn om weg te gaan bij de organisatie, ook al zou ik dat willen' in deze schaal goed weggelaten kunnen worden.

De schaal 'affectief commitment naar het beroep toe' meet vooral goed voor mensen met een lage θ (mensen met een negatief affectief commitment naar hun beroep toe), terwijl deze gezien het antwoordpatroon juist een goed onderscheid zou moeten maken bij de mensen met een positief affectief commitment naar hun beroep toe (dus mensen met een hoge θ). Toch is de schaal redelijk betrouwbaar als de items 5 en 6 weggelaten worden, zeker ook omdat deze twee items ook na de variantie analyse niet veel extra informatie opleveren.

De schaal 'normatief commitment naar het beroep toe' meet betrouwbaar en meet over een groot deel van het theta continuüm. Alleen voor mensen met een sterk negatief normatief commitment naar hun beroep toe, meet deze schaal minder precies.
Vraag 3 'ik voel geen enkele verplichting om dit beroep te blijven uitoefenen' geeft de minste informatie in deze schaal. Deze vraag geeft weinig inzicht in de verschillende antwoordpatronen van de respondenten en kan dus, zeker gezien het feit dat weglating van deze vraag ook voor verhoging van de betrouwbaarheid zorgt, goed weggelaten worden.

De schaal 'continue commitment naar het beroep toe' geeft een hoge betrouwbaarheid. Hij meet goed in het gehele gebied van het theta continuüm, alleen voor de heel hoge en heel lage waarden van θ (mensen met een heel negatieve dan wel heel positieve houding ten opzichte van hun beroep) meet hij minder precies.
De minst informatieve vraag uit deze schaal 'Ik voel me gedwongen om dit beroep uit te oefenen' levert echter wel informatie op over het verschil in antwoordpatroon tussen respondenten van verschillende opleidingsniveau's, verschillende leeftijd, verschillende werktijd in jaren en de verschillende sectoren waarin ze werkzaam zijn. Omdat weglating van de vraag nauwelijks voor verhoging van de betrouwbaarheid zorgt, kan de vraag in de vragenlijst blijven.

De laatst onderzochte schaal is een schaal die uit drie soorten 'fit' bestaat. Uit de gegevens blijkt dat je de 3 subschaaltjes uit de schaal 'hoe vindt u dat u en de organisatie bij elkaar passen' zeer goed kunt samenvoegen tot 1 schaal met de naam 'fit '. De schaal is zeer betrouwbaar –de betrouwbaarste van alle 9 schalen- en meet over een groot deel van het theta continuüm, dus voor zowel mensen die vinden dat ze helemaal niet passen binnen hun organisatie/baan als voor mensen die vinden dat ze wel passen binnen hun organisatie/baan. De minst informatieve items voor de schaal 'fit' zijn de items 1,5 en 9 , maar dit zijn alle vragen over persoon-organisatie fit, dus de vraag is of het wenselijk is om al deze vragen te verwijderen. Duidelijk is in deze schaal te zien dat er steeds 3 keer dezelfde vraag is gesteld (in andere bewoording) in de categorieën persoon-organisatie fit, persoon-baan fit en de behoefte voorziening. Om de vragenlijst te verkorten zou dus beter gekozen kunnen worden om van elke categorie het minst informatieve item te verwijderen, te weten resp. de vragen 2, 5 en 6.

We kunnen dus zeggen dat ondanks dat niet alle schalen in een even breed gebied van theta meten, alle 9 schalen wel goed bruikbaar zijn en zowel een hoge marginale betrouwbaarheid als een hoge Cronbach's alpha hebben. Dat wil zeggen dat binnen elke afzonderlijke schaal de vragen betrekking hebben op dezelfde latente vaardigheid. De betrouwbaarheid is een ratio van de variatie in werkelijk gegeven responsen tegen de variatie van de verwachte responsen (inclusief meetfout). Als de meetfout klein is, zal de ratio dicht bij 1 liggen en is de test betrouwbaar te noemen. De maten van betrouwbaarheid staan in onderstaande tabel nog eens overzichtelijk naast elkaar:

Schaal	Marginale betrouwbaarheid	Cronbach's alpha
1) Vaardigheden van de organisatie	0,860	0,832
2) Uw vaardigheden	0,822	0,785
3) Affectief commitment (organisatie)	0,900	0,842
4) Normatief commitment (organisatie)	0,776	0,727
5) Continue commitment (organisatie)	0,903	0,835
6) Affectief commitment (beroep)	0,822	0,734 (0,789 bij weglating van vraag 5 en 6)
7) Normatief commitment (beroep)	0,858	0,806 (0,814 bij weglating van vraag 3)
8) Continue commitment (beroep)	0,890	0,822
9) 'Fit' – passendheid	0,919	0,892

De toepassing van de Item Respons Theorie op de vragenlijst roept de vraag op of deze meer informatie geeft dan de klassieke analyse. Puur kijkend naar de uitkomsten van de verschillende schalen, geven beide methoden hetzelfde antwoord: de schalen zijn (redelijk) betrouwbaar, met dit verschil dat de betrouwbaardheidswaarden van de IRT analyse consequent iets hoger uitvallen. De IRT analyse heeft als voordeel dat elk item uit elke schaal apart geanalyseerd wordt en dat dit visueel inzichtelijk gemaakt wordt middels de item karakteristieke- en item informatie curves. De IRT analyse levert dus niet zo zeer meer of betere inzichten in de vragenlijst als wel meer inzicht in de individuele vragen uit de schaal alsook meer inzicht in welk gebied van de latente vaardigheid goed wordt gemeten. Een goed voorbeeld hiervan is de schaal 'uw vaardigheden' uit de analyse van deze schaal blijkt dat de schaal niet nauwkeurig meet bij de hogere latente vaardigheden, terwijl de vragen in deze schaal juist veelvuldig in de hogere categorieën geantwoord wordt en hij juist daar wel goed zou moeten meten.

In deze scriptie is de analyse van de schalen uitgevoerd door middel van een unidimensioneel IRT model, die de correlaties tussen de latente vaardigheden negeert. Multidimensionele IRT modellen gebruiken de correlatie tussen de latente vaardigheden wel om de meetprecisie van de individuele latente vaardigheden te

verbeteren en zo te zorgen voor een hogere betrouwbaarheid of een even grote betrouwbaarheid door middel van minder items. De omvang van de verbetering hangt af van de correlatie tussen latente vaardigheden, het aantal subschalen en de lengte van de test. Hoe hoger de correlaties, hoe groter het aantal subschalen en hoe korter de subschalen zijn, des te significanter zal de verbetering zijn. Ook kan er extra achtergrond informatie geïntegreerd worden in de multidimensionele benadering, zoals geslacht, opleiding en leeftijd. Voor dit laatste zijn meer studies nodig om de gevolgen en toepassingen van het toevoegen van achtergrondvariabelen te onderzoeken (Wang, Chen & Cheng, 2004). Bij vervolgonderzoek zou de vragenlijst met behulp van een multidimensionele benadering gedaan kunnen worden, waarbij dan ook de achtergrondvariabelen meegenomen zouden kunnen worden. Omdat uit de variantie analyse van het minst informatieve item per schaal nu ook al blijkt dat er (grote) verschillen zijn in antwoordpatroon tussen respondenten van bijvoorbeeld verschillende leeftijden of opleidingen is het interessant om deze factoren in samenhang met de vragenlijst nader te onderzoeken.

Referentielijst

Bakker, F.B. (2001). *The Basics of Item Response Theory*. ERIC Clearinghouse on Assesment and evaluation.

Bakker, F.B.& Kim,S. (2004). *Item Response Theory: Parameter Estimation Techniques*. New York: Marcel Dekker Inc.

Cable, D.M. (2002). The Convergent and Discriminant Validity of Subjective Fit Perceptions. *Journal of Applied Psychology*, 87 (5), 875-884.

Cable, D.M., & Judge, T.A. (1996). Person-organization fit, job choice decisions, and Organizational entry. *Organizational Behvior and Human Decision Processes*, 67, 294-311.

Cable, D.M., & Judge, T.A. (1997). Interviewers' perceptions of person-organizational Fit and organizational selection decisions. *Journal of Applied Psychology*, 82, 546-561.

Eisenberger, R., Huntington, R., Hutchinson, S., & Sowa, D. (1986). Perceived Organizational support. *Journal of Applied Psychology*, 71, 500-507.

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

Flannery, W. P., Reise, S. P.,& Widaman, K. F. (1995). An item response theory analysis of the general and academic scales of the self-description questionnaire II. *Journal of Research in Personality*, *29*, 168-188.

Gibson, C.B., Randel, A.E., & Earley, P.C. (2000). Understanding group efficacy: An emperical test of multiple assessment methods. *Group & Organization Management*, 25, 67-97.

Hambleton, R.K.,Swaminathan, H.,& Rogers, H.J. (1991) *Fundamentals of Item Response Theory*. Newburry Park,Calif.: Sage Publications.

Hambleton,R.K.,& Swaminathan,H. (1985). *Item Response Theory: Principles and Applications*. Boston,MA:Kluwer-Nijhoff.

Judge, T.A., & Cable, D.M. (1997). Applicant personality, organizational culture, and Organization attraction. *Personnel Psychology*, 50, 359-393.

Kaliath, T.J., Bluedorn, A.C., & Strube, M.J. (1999). A test of value congruence effects.*Journal of Organizational Behavior*, 20, 1175-1198.

Kristof-Brown, A.L. (2000). Perceived applicant fit: Distinguishing between recruiters' Perceptions of person-job and person-organization fit. *Personnel Psychology*, 53, 643-671.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Masters, G. N.,& Wright, B. D. (1997). The partial credit model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 101-121). New York, NY: Springer-Verlag Inc.

Mellenbergh, G. J. (1994). A unidimensional latent trait model for continuous item responses. *Multivariate Behavioral Research, 29*, 223-236.

Meyer, J.P., & Allen, N.J. (1991). A three-component conceptualization of organizational Commitment. *Human Resource Management Review*, 1, 61-98.

Meyer, J.P., Allen, N.J., & Smith, C.A. (1993). Commitment to Organizations and Occupations: Extensions and Test of a Three-Component Conceptualization. *Journal Of Apllied Psychology*, 4, 538-551.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 17, 351-363.

Mowday, R.T., Porter, L.W., & Steers, R.M. (1982). *Employee-organizational linkages:The psychology of commitment, absenteeism, and turnover*. San Diego, CA: Academic Press.

Ostini, R., & Nering, M.L. (2005).*Polytomous Item Response Theory Models*. Thousand Oaks, Calif.: Sage Publications.

Saks, A.M., & Ashforth, B.E. (1997). A longitudinal investigation of the relationship Between job information sources, applicant perceptions of fit, and work outcomes. *Personnel Psychology*, 50, 395-426.

Shamir, B.,House, R.J., & Arthur, M.B. (1993). The motivational effects of charismatic leadership: A self-concept based theory. *Organization Science*, 4, 1-17.

Thissen, D. (1986). MULTILOG: Item analysis and scoring with multiple category response models (Version 6). Mooresville, IN: Scientific Software.

Verhelst, N.D. (1992).*Het eenparameter logistisch model (OPLM): Een theoretische inleiding en een handleiding bij het computerprogramma*. Arnhem: Cito.

Vuuren, M. Van, Elving, W.J.L.,de Jong M.D.T., & Seydel, E.R. (2005). 'What keeps you here?' : Organizational efficacy and person-organization fit as antecedents of affective organizational commitment.

Wang, W., Chen, P.,& Cheng,Y. (2004). Improving Measurement Precision of Test Batteries Using Multidimensional Item Respons Models. Psychological Methods *2004*, 9 (1),116-136.

Weiss, D. J. (1995). Improving individual differences measurement with item response theory and computerized adaptive testing. In D. J. Lubinski, & R. V. Dawis (Eds.), *Assessing individual differences in human behavior: New concepts, methods, and findings* (pp. 49-79). Palo Alto, CA: Davies-Black Publishing.

Bijlage variantie analyse

Vaardigheden van de organisatie – item 2

	F-waarde	Significantie waarde
Geslacht	0,070	0,792
Opleiding	1,237	0,289
Leeftijd in jaren	1,040	0,392
Sector waar werkzaam	17,696	0,000
Jaren werkzaam in organisatie	1,788	0,112

Uw Vaardigheden – item7

	F-waarde	Significantie waarde
Geslacht	0,101	0,751
Opleiding	5,035	0,000
Leeftijd in jaren	3,905	0,002
Sector waar werkzaam	5,811	0,001
Jaren werkzaam in organisatie	2,385	0,036

Uw houding ten opzichte van uw organisatie – affectief commitment – item 5

	F-waarde	Significantie waarde
Geslacht	0,016	0,898
Opleiding	4,437	0,001
Leeftijd in jaren	14,491	0,000
Sector waar werkzaam	67,935	0,000
Jaren werkzaam in organisatie	10,602	0,000

Uw houding ten opzichte van uw organisatie – normatief commitment – item 5

	F-waarde	Significantie waarde
Geslacht	9,185	0,002
Opleiding	1,575	0,164
Leeftijd in jaren	18,307	0,000
Sector waar werkzaam	18,872	0,000
Jaren werkzaam in organisatie	5,709	0,000

Uw houding ten opzichte van uw organisatie – continue commitment – item 1

	F-waarde	Significantie waarde
Geslacht	0,223	0,637
Opleiding	10,602	0,000
Leeftijd in jaren	23,921	0,000
Sector waar werkzaam	12,766	0,000
Jaren werkzaam in organisatie	26,241	0,000

Uw houding ten opzichte van uw beroep – affectief commitment – item 5

	F-waarde	Significantie waarde
Geslacht	0,203	0,652
Opleiding	1,695	0,133
Leeftijd in jaren	1,750	0,120
Sector waar werkzaam	5,349	0,001
Jaren werkzaam in organisatie	1,171	0,321

Uw houding ten opzichte van uw beroep – affectief commitment – item 6

	F-waarde	Significantie waarde
Geslacht	0,001	0,972
Opleiding	3,596	0,003
Leeftijd in jaren	2,280	0,045
Sector waar werkzaam	6,403	0,000
Jaren werkzaam in organisatie	1,692	0,134

Uw houding ten opzichte van uw beroep – normatief commitment – item 3

	F-waarde	Significantie waarde
Geslacht	0,815	0,367
Opleiding	0,431	0,827
Leeftijd in jaren	4,939	0,000
Sector waar werkzaam	3,061	0,027
Jaren werkzaam in organisatie	2,668	0,021

Uw houding ten opzichte van uw beroep – continue commitment – item 4

	F-waarde	Significantie waarde
Geslacht	2,272	0,132
Opleiding	6,139	0,000
Leeftijd in jaren	14,556	0,000
Sector waar werkzaam	22,632	0,000
Jaren werkzaam in organisatie	14,362	0,000

Hoe u vindt dat de organisatie en u bij elkaar passen – item 2

	F-waarde	Significantie waarde
Geslacht	0,195	0,659
Opleiding	1,708	0,130
Leeftijd in jaren	1,857	0,099
Sector waar werkzaam	23,389	0,000
Jaren werkzaam in organisatie	2,921	0,013

Hoe u vindt dat de organisatie en u bij elkaar passen – item 5

	F-waarde	Significantie waarde
Geslacht	0,013	0,908
Opleiding	3,250	0,006
Leeftijd in jaren	2,557	0,026
Sector waar werkzaam	33,868	0,000
Jaren werkzaam in organisatie	4,809	0,000

Hoe u vindt dat de organisatie en u bij elkaar passen –item 6

	F-waarde	Significantie waarde
Geslacht	0,461	0,497
Opleiding	2,404	0,035
Leeftijd in jaren	6,112	0,000
Sector waar werkzaam	2,288	0,077
Jaren werkzaam in organisatie	1,177	0,318

Verklaring van de legenda bij de histogrammen:

Leeftijd

1- jonger dan 20 jaar
2- 20 t/m 30 jaar
3- 31 t/m 40 jaar
4- 41 t/m 50 jaar
5- 51 t/m 60 jaar
6- 61 jaar en ouder

Jaren werkzaam bij de organisatie

1- korter dan 1 jaar
2- 1 tot 2 jaar
3- 2 tot 5 jaar
4- 6 tot 10 jaar
5- 11 tot 15 jaar
6- langer dan 15 jaar

Opleiding

1- lager onderwijs
2- middelbaar onderwijs
3- middelbaar beroepsonderwijs
4- hoger beroepsonderwijs
5- wetenschappelijk onderwijs
6- anders

Sector

1- personeel uit een zorginstelling
2- vrijwilligerspersoneel uit een zorginstelling
3- personeel werkzaam in een petro chemische fabriek
4- personeel uit een ziekenhuis

Bijlage output SPSS

Reliability Statistics

Cronbach's Alpha	N of Items
,832	7

Item Statistics

	Mean	Std. Deviation	N
ce1	3,92	1,661	1387
ce2	4,61	1,379	1387
ce3	4,11	1,449	1387
ce4	4,24	1,432	1387
ce5	5,12	1,350	1387
ce6	4,52	1,479	1387
ce7	4,59	1,504	1387

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
ce1	27,19	37,388	,610	,804
ce2	26,50	42,765	,436	,830
ce3	27,00	39,332	,610	,804
ce4	26,87	38,524	,672	,794
ce5	25,99	41,735	,514	,819
ce6	26,59	39,252	,598	,806
ce7	26,52	38,702	,618	,803

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
31,11	52,523	7,247	7