



UvA-DARE (Digital Academic Repository)

Testing distributional assumptions in psychometric measurement models with substantive applications in psychology

Molenaar, D.

Publication date
2012

[Link to publication](#)

Citation for published version (APA):

Molenaar, D. (2012). *Testing distributional assumptions in psychometric measurement models with substantive applications in psychology*.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Dutch Summary

Het testen van verdelingsaannamen in psychometrische meetmodellen met inhoudelijke toepassingen in de psychologie

In de psychologie zijn de constructen die onderzocht worden over het algemeen niet direct te observeren. Daarom wordt er gesproken van *latente variabelen*. Om uitspraken te kunnen doen over deze variabelen wordt er gekeken naar observeerbare indicatoren van de latente variabelen. Bijvoorbeeld in het geval van de latente variabele ‘ruimtelijk inzicht’, kan er gekeken worden naar de (observeerbare) prestaties van een groep proefpersonen op de Blok Patronen en Matrix Redeneren subschalen van de Wechsler Adult Intelligence Scales (WAIS).

In een *psychometrisch meetmodel* wordt het verband tussen de geobserveerde variabelen en de latente variabelen beschreven. Er zijn verschillende families van meetmodellen. Het geschikte meetmodel hangt af van de aard van de geobserveerde en latente variabelen. Een overkoepelend meetmodel - dat veel populaire meetmodellen omvat - is het Gegeneraliseerde Lineaire Item Respons Model (GLIRM). Voor alle modellen binnen het GLIRM geldt dat -door de afhankelijke variabelen te transformeren- alle meetmodellen te herleiden zijn tot een lineair model. De parameters van dit lineaire meetmodel zijn: Regressiecoëfficiënten (factorladingen), intercepts, en residuele varianties.

Dit proefschrift focust op 2 speciale meetmodellen binnen het GLIRM: 1) Het lineaire factor model, omdat dit model veelvuldig gebruikt wordt om subtest scores te modeleren (bijv. van intelligentie tests), en 2) Het ‘graded response model’, omdat dit model een geschikt model is om variabelen met Likert schalen te modeleren (bijv. van persoonlijkheidstests).

Wanneer een juist meetmodel is gekozen, kunnen hypothesen over de latente variabelen getoetst worden in het structurele model. Hiertoe dient eerst een verdeling voor de getransformeerde data gespecificeerd te worden. In zowel het lineaire factor model als het graded response model wordt overwegend een normale verdeling gebruikt. Dit proefschrift presenteert modellen die gericht zijn op het toetsen van deze aanname. Er zijn twee redenen waarom deze modellen als waardevol gezien kunnen worden. Ten eerste bestaan er geen specifieke statistische toetsen op deze aanname. Het is waardevol om zulke toetsen beschikbaar te hebben om te onderzoeken of het aannemen van een normale verdeling gerechtvaardigd is. Verder kan er met zulke modellen onderzocht worden wat de consequenties zijn van schendingen van normaliteit in een bepaalde dataset. In de literatuur zijn gevallen gerapporteerd waarin

het ten onrechte aannemen van een normale verdeling voor de getransformeerde data tot vertekeningen van de statistische resultaten leidde. Ten tweede zijn de modellen die voorgesteld worden in dit proefschrift waardevol als toetsen op inhoudelijke hypothesen zoals *ability differentiation*, *schematiciteit*, en *genotype bij omgevingsinteracties*. Al het bovenstaande wordt besproken in het inleidende *Hoofdstuk 1*.

In *Hoofdstuk 2* wordt beargumenteerd dat binnen het eendimensionale lineaire factor model, de aanname van een normale verdeling voor de data verworpen moet worden als: 1) de latente variabele niet normaal verdeeld is; 2) de factorladingen afhangen van de latente variabele; en/of 3) de residuele varianties heteroscedastisch zijn. In dit Hoofdstuk worden deze drie effecten gemodelleerd binnen het lineaire factormodel. Hierbij wordt gebruik gemaakt van:

- #1: een scheefnormale verdeling voor de latente variabele;
- #2: een parametrische functie tussen de factor ladingen en de latente variabele;
- #3: een parametrische functie tussen de residuele varianties en de latente variabele.

Er wordt aangetoond dat een lineair factor model met één van de genoemde effecten afzonderlijk succesvol kan worden toegepast op data. Verder wordt er aangetoond dat effect #1 en #3 gecombineerd kunnen worden in één model, en dat #2 en #3 kunnen worden gecombineerd in één model. Bij realistische steekproefgroottes blijkt het echter niet mogelijk om #1 met #2 te combineren. Het Hoofdstuk sluit af met een toepassing van de verschillende modellen op een intelligentie dataset om te toetsen op *ability differentiation*. Deze hypothese houdt in dat de algemene intelligentie factor, g , niet een even sterke bron van individuele verschillen is op verschillende punten van de g schaal. Dit zou de aanwezigheid van #1, #2, en/of #3 impliceren. Voor deze hypothese wordt echter geen ondersteuning gevonden.

In *Hoofdstuk 3* worden binnen het graded response model voor ordinale data de effecten #1 en #3 gecombineerd in één model. Dit model wordt toegepast op persoonlijkheidsdata om de schematiciteitshypothese te toetsen. Deze hypothese houdt in dat mensen met een laag niveau op een bepaalde persoonlijkheidsvariabele onzekerder zijn over hun exacte positie op die variabele. Dit zou impliceren dat voor een gegeven persoonlijkheidsvariabele, de residuele varianties van de verschillende persoonlijkheidsvragen afnemen langs de latente variabele. In het Hoofdstuk wordt ondersteuning gevonden voor deze hypothese in een alexithymia dataset. Verder wordt er gevonden dat de latente variabelen -die volgens de alexithymia theorie ten grondslag liggen aan deze data- negatief scheef verdeeld zijn.

In *Hoofdstuk 4* wordt de methodologie voor het eendimensionale lineaire factormodel uitgebreid naar een model met meerdere eerste-orde factoren en één

tweede-orde factor. Deze uitbreiding wordt gepresenteerd met als doel te toetsen op ability differentiation. Hiertoe wordt effect #1 toegepast op de tweede-orde factor, en effect #3 wordt toegepast op de residuele varianties van de geobserveerde subtest scores in het model. In een toepassing van het model op de Spaanse standaardisatiedata van de WAIS wordt een negatief scheve g verdeling gevonden. Dit resultaat ondersteunt de ability differentiation hypothese. Verder worden er enkele heteroscedastische residuele varianties gevonden.

In *Hoofdstuk 5* wordt de methode van gemodereerde factor analyse gepresenteerd. In deze methode worden de parameters van het traditionele tweede-orde factor model (zie Hoofdstuk 4) gemodereerd door een geobserveerde variabele (de moderator) zoals leeftijd. Als voor de moderator een proxy voor de tweede-orde factor wordt gebruikt kan deze methode gebruikt worden om te toetsen op #1, #2, en/of #3. In dit Hoofdstuk wordt de methode gebruikt om te toetsen op *age differentiation* en ability differentiation. Age differentiation houdt in dat g niet een even sterke bron van individuele verschillen is voor verschillende leeftijden. In een tweelingendataset wordt aangetoond dat de traditionele manier van principale componenten analyse ongeschikt is om deze twee hypothesen te onderzoeken omdat er geen verschil wordt gemaakt tussen de g factor en residuele factoren. Bij het toepassen van het gemodereerde tweede-orde factormodel wordt geen ondersteuning gevonden voor age differentiation en ability differentiation.

In *Hoofdstuk 6* wordt het lineaire factor model met effect #3 uitgebreid naar het klassieke tweelingendesign. Dit wordt gedaan om te kunnen toetsen op een genotype bij omgevinginteractie (GxE). Een GxE impliceert dat de omgevingsfactor, E, niet constant is langs de additief genetische factor, A. Dus, een mogelijke toets op GxE is een toets op homoscedasticiteit van E. In dit Hoofdstuk wordt zowel een univariate methode gepresenteerd als een multivariate methode. In de multivariate methode wordt de GxE expliciet gemodelleerd op het niveau van de latente variabele. Hierbij wordt op het niveau van de geobserveerde variabelen mogelijke heteroscedasticiteit in beschouwing genomen. Dit heeft het voordeel dat scheefheid van de geobserveerde data door andere oorzaken (bijvoorbeeld door bodem effecten, plafond effecten, of slechte schaling van de geobserveerde variabelen) minder snel zal leiden tot vertekende resultaten. In een toepassing van het multivariate model op een intelligentie dataset wordt een GxE gevonden op g .

In *Hoofdstuk 7* wordt het univariate model uit het vorige Hoofdstuk toegepast op een grote tweelingendataset. Deze data omvat intelligentiescores van tweelingen uit 14 verschillende landen. De resultaten zijn zeer heterogeen, onder andere met betrekking tot het GxE effect. Dit suggereert dat de richting en grootte van het GxE effect verschilt tussen verschillende landen.

In *Hoofdstuk 8* wordt besproken wat de consequentie is van het verbinden van een inhoudelijke interpretatie aan een statistisch effect. Dit moet met de nodige voorzichtigheid gedaan worden omdat de statistische effecten alternatieve verklaringen kunnen hebben. Er worden mogelijke storende factoren besproken en richtlijnen voor het verbinden van een inhoudelijke conclusie aan een statistisch effect. Ook worden er verdere toepassingen en ontwikkelingen besproken voor de modellen in dit proefschrift.

In *Hoofdstuk 9* wordt besproken hoe binnen het GLIRM power berekeningen gedaan kunnen worden voor de *likelihood ratio test*. Nagenoeg alle Hoofdstukken in dit proefschrift gebruiken deze methode. De verhandeling wordt geïllustreerd met een voorbeeld over sekseverschillen in intelligentietestscores.