

University of Groningen

## Dimensionality assesment with factor analysis methods

Barendse, Mariska

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2015

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Barendse, M. (2015). *Dimensionality assesment with factor analysis methods*. [Thesis fully internal (DIV), University of Groningen]. [S.n.].

### Copyright

Other than for strictly personal use, 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), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

## Summary in Dutch

In dit proefschrift staat de dimensionaliteitsbepaling van itemscores centraal. De formele grondslag van dimensionaliteit wordt gevormd door definities van Lord & Novick (1968) en Zhang & Stout (1999). In deze definities wordt dimensionaliteit weergegeven als het minimale aantal latente trekken dat nodig is om de statistische afhankelijkheid in de data volledig te beschrijven. Vanuit praktisch oogpunt is het bepalen van de dimensionaliteit belangrijk voor de constructie en de evaluatie van vragenlijsten. Deze dimensionaliteit geeft inzicht in de structuur die hieraan ten grondslag ligt aan de itemscores. Er zijn veel methoden om de dimensionaliteit te bepalen, waarbij de factoranalytische methoden het meest gebruikt worden (zie Conway & Huffcutt, 2003; Ten Holt en anderen, 2010).

Dit proefschrift richtte zich op factoranalytische (FA) methoden om de dimensionaliteit te bepalen. De belangrijkste assumptie hierbij is dat de dimensionaliteit van een dataset gelijk is aan het minimum aantal gemeenschappelijke factoren. In FA meetmodellen wordt het verband tussen een groot aantal geobserveerde items door een kleiner aantal latente variabelen, de factoren, beschreven. De geobserveerde variantie van elke variabele (item) kan opgesplitst worden in een gemeenschappelijk deel en een uniek deel. Het gemeenschappelijke deel beschrijft wat een geobserveerde variabele gemeen heeft met de andere geobserveerde variabelen. Hierbij geven de elementen van de zogenaamde patroonmatrix het verband weer tussen de geobserveerde variabelen en de latente variabelen. Het unieke deel is niet (lineair) gerelateerd aan de geobserveerde variabelen en bestaat uit een itemspecifiek deel en een meetfout deel (zie Bollen, 1989).

De klassieke factoranalytische methoden zijn geschikt voor het beschrijven van lineaire relaties tussen multivariaat normaal verdeelde itemscores. In de praktijk zijn itemscores vaak polytoom. Bovendien is er soms een hiërarchische (multilevel) structuur aanwezig. Het doel van dit proefschrift was om te onderzoeken of factoranalytische methoden geschikt zijn om de dimensionaliteit te bepalen bij complexe datastructuren.

Ik richtte me in het bijzonder op exploratieve factoranalyse (**Hoofdstuk 2 en 3**), multilevel data (**Hoofdstuk 3**), vraagonzuiverheid (**Hoofdstuk 4 en 5**) en op het analyseren van polytome items (**Hoofdstuk 2, 3, en 4**). Het proefschrift bestaat uit op zichzelf staande hoofdstukken, voorafgegaan door een introductie en afgesloten door een samenvatting en een discussie.

In **hoofdstuk 2** behandelden we het gebruik van exploratieve factoranalyse (EFA) om de dimensionaliteit te bepalen van polytome itemscores. In een EFA worden geen restricties opgelegd aan de patroonmatrix (behalve ter identificatie). In EFA kunnen verschillende fitstatistieken gebruikt worden om het aantal factoren te bepalen. Bij continue data worden itemparameters geschat door factoranalyse toe te passen op de variantie-covariantiematrix van de itemscores. Er zijn verschillende studies waarin de dimensionaliteit bij continue data is onderzocht (bijv. Conway & Huffcutt, 2003; Fabrigar et al., 1999; Preacher & MacCallum, 2003). In de praktijk bestaan veel testen uit polytome itemscores (bijv. “correct” / “niet correct” of Likert schalen die variëren van “helemaal van toepassing” tot “niet van toepassing”), hetgeen het onderzoek naar de dimensionaliteitsbepaling bij polytome variabelen van belang maakt. Onderzoek heeft uitgewezen dat het behandelen van polytome itemscores op dezelfde manier als continue itemscores zorgt voor afwijkende parameterschattingen, incorrecte standaard fouten en incorrecte fitstatistieken. Bij het analyseren van polytome variabelen wordt dan ook de polychorische correlatiematrix, in plaats van de covariantiematrix, geanalyseerd. Er zijn tot nu toe weinig studies die onderzoek doen naar het bepalen van dimensionaliteit bij discrete data (zie Tate, 2003; Holgado-Tello et al., 2010; Gaskin & Happell, 2014).

We beschreven in **hoofdstuk 2** een simulatiestudie om de dimensionaliteit van discrete data te bepalen. We varieerden veel voorkomende data karakteristieken, zoals de grootte van factorladingen, het aantal antwoorδοpties, de steekproefgrootte en de complexiteit van de factorstructuur. De resultaten van het gebruik van verschillende schattingsmethoden (ML, robuuste ML, ML van polychorische correlaties, weigthd least squares

(WLS), robuuste WLS) wezen uit dat de robuuste WLS methode de meest efficiënte schattingsmethode is voor polytome variabelen en dat deze ook de minste convergentieproblemen gaf. De simulatiestudie liet zien dat het erg moeilijk was om zowel de grote als de kleine factoren met exacte fitstatistieken te identificeren. Fitstatistieken die gebaseerd zijn op residuen en de meer globalere fitstatistieken (root mean square error of approximation) bleken geschikt om de grote factoren te identificeren. Bij alle schattingsmethoden kwamen we convergentieproblemen tegen, vooral als er veel factoren zijn en als de steekproeven klein zijn. In de praktijk, als we bij de bepaling van de dimensionaliteit van een enkele dataset convergentieproblemen zouden tegenkomen, zijn mogelijke oplossingen om verschillende identificatierestricties en startwaarden te proberen. Het zou interessant zijn om de convergentie van andere schattingsmethoden dan die in Hoofdstuk 2 te onderzoeken.

In **hoofdstuk 3** onderzochten we de dimensionaliteitsbepaling van discrete data met een multilevel datastructuur. Bij psychologische of onderwijskundige vragenlijsten komen data met multilevel datastructuren veel voor. In dergelijk onderzoek worden dan bijvoorbeeld scholieren uit verschillende schoolklassen of kinderen uit verschillende families onderzocht. In dit hoofdstuk wordt de voorgestelde aanpak geïllustreerd met data van een onderwijskundige vragenlijst afgenomen onder kinderen uit verschillende schoolklassen. Om de dimensionaliteit op verschillende niveaus (“levels”) te bepalen, maakten we gebruik van een EFA en de robuuste WLS schattingsmethode. We vergeleken een procedure zonder restricties over verschillende niveaus met een procedure die restricties over verschillende niveau oplegt. De procedure zonder restricties over verschillende niveaus laat ruimte vrij voor een verschillend aantal factoren op iedere niveau en in deze procedure hebben de factoren op verschillende niveaus niet vanzelfsprekend dezelfde betekenis. De procedure met restricties veronderstelde een gelijk aantal factoren op elk niveau die ook dezelfde betekenis hebben over verschillende niveaus. De resultaten van de empirische data analyse liet zien dat de procedure met restricties over de niveaus de best interpreteerbare uitkomsten gaf.

In **hoofdstuk 4** bespraken we het gebruik van de relatief nieuwe

“pairwise maximum likelihood” (PML; Jöreskog & Moustaki, 2001) methode die aangewend kan worden om discrete data te analyseren. Omdat er nog geen fitstatistieken algemeen beschikbaar zijn, introduceerden we drie nieuwe fitstatistieken. De nieuwe fitstatistieken zijn gebaseerd op de likelihood ratios, namelijk  $C_F$  (vergelijkt de model geïmpliceerde proporties met de geobserveerde proporties van volledige antwoordpatronen),  $C_M$  (vergelijkt de model geïmpliceerde proporties van volledige antwoordpatronen met de verwachte proporties onder de multivariate normaliteitsassumptie) en de  $C_P$  (vergelijkt de model geïmpliceerde proporties van paarsgewijze antwoordpatronen onder de bivariate verdeling met de geobserveerde proporties van paarsgewijze antwoordpatronen). De simulatiestudie met verschillende steekproefgroottes en een verschillend aantal antwoordmogelijkheden wees uit, dat de parameterschattingen van de PML en de robuuste WLS bijna identiek waren. Daarnaast bleken de  $C_M$  en  $C_P$  goed bruikbaar voor model selectie. Hoewel de PML niet beter presteerde dan de robuuste WLS, kan de PML toch gezien worden als een veelbelovende schattingsmethode, omdat er in tegenstelling tot de robuuste WLS sprake is van een éénstaps schattingsmethode. Vanuit theoretisch oogpunt is dit beter. Wellicht dat bij grote datasets en “missing data” (d.w.z. niet ingevulde vragen) de PML te prefereren is boven de robuuste WLS.

In **hoofdstuk 5** en **6** richtten we ons zich op vraagonzuiverheid, hetgeen nauw gerelateerd is aan dimensionaliteit. Bij vraagonzuiverheid wordt er onderzocht of geobserveerde verschillen in itemscores een valide representatie is van eigenlijke verschillen in scores op de latente variabelen (Oort, 1996 naar Mellenbergh, 1989). Vanuit een modelmatig perspectief is een vraagonzuivere variabele een mogelijke extra dimensie in een FA meetmodel. Vraagonzuiverheid in een één-factormodel impliceert dus multidimensionaliteit. Als multidimensionaliteit niet correct is gemodelleerd kan het resulteren in vraagonzuiverheid (zie Meredith, 1993; Jak et al., 2010).

Vraagonzuiverheid ten opzichte van verschillende groepen kan onderzocht worden met een multigroep factoranalyse (MGFA; Meredith, 1993). Daarnaast kan vraagonzuiverheid ten opzichte van groepslidmaatschap

of elke andere variabele die vraagonzuiverheid kan veroorzaken (de zogenaamde “schender”) ook onderzocht worden met een restrictieve factoranalyse (RFA; Oort, 1992, 1998). De RFA methode onderzoekt vraagonzuiverheid door middel van een exogene variabele. Voordelen van het RFA model zijn dat de schender continue of discreet, maar ook geobserveerd of latent kan zijn; daarnaast kan vraagonzuiverheid gedetecteerd worden ten opzichte van meerdere variabelen. Uniforme vraagonzuiverheid wordt dan geïndiceerd door een direct effect van de schender op de itemscore (Oort, 1992, 1998). Niet-uniforme vraagonzuiverheid wordt geïndiceerd door een interactie effect van de schender met het gemeten construct op de itemscore (Barendse et al., 2010).

In **hoofdstuk 5** evalueerden we de RFA methode in een simulatiestudie en vergeleken we de resultaten met die van de MGFA methode. Niet-uniforme bias wordt door middel van twee verschillende methoden geoperationaliseerd, namelijk de “random slope parametrization” (Muthén & Asparouhov, 2003) en de “latent moderated structures” (Klein & Moosbrugger, 2000). Vraagonzuiverheid was geïndiceerd bij verschillende significantieniveaus en in verschillende vraagonzuiverheidsprocedures (niet-iteratief, iteratief). Verder varieerden we het type vraagonzuiverheid (uniform, niet-uniform, beide, geen van beide), de afhankelijkheid tussen schender en het gemeten construct (afhankelijkheid, onafhankelijkheid) en het type schender (continu, discreet). Resultaten wezen uit dat de parameterschattingen van zowel de MGFA als de RFA efficiënt en accuraat zijn. Met een continue schender zijn de resultaten van vraagonzuiverheid bij de RFA methode veel beter dan die bij de MGFA methode. Uit de resultaten concludeerden we verder dat de iteratieve vraagonzuiverheid-detectiemethode de voorkeur heeft boven een niet-iteratieve methode.

In **hoofdstuk 6** van dit proefschrift beschreven we een Bayesiaanse aanpak om zowel uniforme als niet-uniforme vraagonzuiverheid te detecteren met de RFA methode bij een continue geobserveerde schender, een continue latente schender en een dichotome schender. Met de aanpak beschreven door Lee (2007) is het mogelijk om de interactie te schatten om niet-uniforme vraagonzuiverheid te detecteren. Uit een simulatieonderzoek met continue items kwam naar voren dat zowel erg infor-

matieve als weinig informatieve priors (parameters voorafgaand aan de analyse die geen vraagonzuiverheid veronderstellen) leiden tot accurate en efficiënte parameterschattingen. Alleen in de condities met een gedichotomiseerde schender waren de parameterschattingen minder efficiënt door het discreet maken van een continue variabele. Voor de detectie van vraagonzuiverheid bleek de verschiltoets (tussen de passing van een model met parameters voor vraagonzuiverheid en de passing van een model zonder parameters om vraagonzuiverheid te detecteren) de beste methode. Hierbij wordt het verschil in “deviance information criterion” gebruikt. Verder concludeerden we dat de convergentie van de geschatte modellen moeilijk bleek bij de Bayesiaanse RFA, zelfs bij het gebruik van convergentie criteria (zie Raftery & Lewis, 1992; Gelman & Rubin, 1992).

In dit proefschrift werden FA methoden gebruikt om de dimensionaliteit te bepalen. Klassieke FA methoden zijn geschikt voor lineaire relaties tussen continue variabelen met een multivariate normale verdeling. We stelden FA methoden en procedures voor die gebruikt kunnen worden om de dimensionaliteit te bepalen in andere situaties. FA methoden bleken in de bestudeerde situaties goed te kunnen omgaan met polytome items, niet-lineaire relaties en multilevel structuren.

Dit proefschrift is gelimiteerd in de zin dat ik me alleen heb gericht op FA methoden en niet op andere klassieke FA methoden, zoals parallel analyse of andere methoden zoals niet-parametrische item response theorie (IRT) gebaseerde methoden. Het zou interessant zijn om de niet-lineaire modellen uit hoofdstuk 5 en 6 te vergelijken met de niet-lineaire IRT modellen beschreven door Molenaar (2012). Daarnaast zouden de modellen uit hoofdstuk 3 vergeleken kunnen worden met de Bayesiaanse of frequentistische IRT georiënteerde multilevel modellen (bijv. Verhagen, 2012).