



Sociaal en Cultureel Planbureau

# Technische bijlage hoofdstuk 7 en 8 Denkend aan Nederland

Sociaal en Cultureel Rapport 2019

Juliette de Wit (RUG), Sjoerd Beugelsdijk (RUG), Sander Muns

Sociaal en Cultureel Planbureau  
Den Haag, juni 2019

## Inhoudsopgave

Hoofdstuk 7 .....	4
Factoranalyse .....	4
1.1 Data.....	4
DAN-enquête.....	4
1.2 Methode.....	6
Confirmatieve factoranalyse .....	9
1.3 Resultaten .....	9
1.3.1 Typerend voor Nederland .....	9
Exploratieve fase .....	9
Confirmatieve fase.....	12
1.3.2 Verbondenheid met de Nederlandse identiteit.....	13
Exploratieve fase .....	13
Confirmatieve fase.....	15
Regressieanalyse.....	16
Lasso-procedure .....	16
Bepaling extreme groepen.....	17
Mahalanobis-afstand .....	17
Hoofdstuk 8 .....	19
2.1 Archetypenanalyse-methode .....	19
2.2 Een typologie van Nederlandse identificatie .....	20
2.3 Robuustheid model.....	22
Conclusie .....	23
Referenties .....	24
Appendix .....	25

## Technische Bijlage Hoofdstuk 7 en 8

Deze bijlage vat de belangrijkste methodologische keuzes van de analyses in de hoofdstukken 7 en 8 samen. De algemene onderzoeksverantwoording van het gehele SCR is opgenomen in hoofdstuk 18. We bespreken in deze bijlage de analyses van de hoofdstukken 7 en 8. In deze bijlage gaan we in op:

- Hoofdstuk 7:
  - Een bespreking van de data, methode en resultaten die aan de basis liggen van de factormodellen die zijn gepresenteerd in hoofdstuk 7:
    - data: een bespreking van de verschillende overwegingen en keuzes voor de items die in de factoranalyses meegenomen zijn;
    - methode: een methodologisch overzicht van exploratieve en confirmatieve factoranalyse;
    - resultaten: een bespreking van de resultaten voor de typerend-items, alsmede voor de verbondenheid-items.
  - Een bespreking van de methodes onderliggend aan de regressieanalyses die gepresenteerd zijn in hoofdstuk 7.
- Hoofdstuk 8:
  - Methode: bespreking van archetypenanalyse.
  - Resultaten: analyse van het aantal archetypen en beslisregels.
  - Robuustheid: analyse van modellen met twee en vier archetypen.
- Conclusie

## Hoofdstuk 7

In deze paragraaf beschrijven we de data die gebruikt zijn in de verschillende analyses, evenals de methodes en de verschillende stappen die gezet zijn om te komen tot de eindresultaten in de verschillende tabellen in hoofdstuk 7.

### Factoranalyse

#### 1.1 Data

##### DAN-enquête

De DAN-enquête is in verschillende fasen ontwikkeld. Dat staat in detail beschreven in hoofdstuk 18. In elke fase zijn twee blokken vragen opgenomen die enerzijds vragen naar dat wat typerend is voor Nederland en anderzijds naar dat waardoor men zich verbonden voelt met Nederland. Zo wordt er bijvoorbeeld gevraagd naar Nederlandse beroemdheden – van Armin van Buuren tot Piet Hein – naar landschapskenmerken – van heide tot de Amsterdamse grachten – en naar Nederlandse gewoonten en tradities – van Koningsdag tot het eten van stamppot. De vragenlijst is in verschillende testrondes verder aangepast; items zijn toegevoegd en verder verfijnd. Voorafgaand aan het uitzetten van de definitieve versie van de DAN-enquête binnen het LISS-panel zijn eerdere versies kwantitatief getest in het I&O-panel. Deze versie bestond onder meer uit 180 typerend- en verbondenheid-items en is voorgelegd aan 1400 respondenten. Er is een aanzienlijke overlap tussen de twee vragenlijsten, maar enkele vragen zijn niet meegenomen in de uiteindelijke versie van de DAN-enquête. Ook zijn er nieuwe items toegevoegd. Tabel 1 geeft een overzicht van deze items. De uiteindelijke versie van de enquête is voorgelegd aan de deelnemers in het LISS-panel.

Tabel 1

Overlap enquêtes I&O en LISS

Items die in de tweede fase eruit gehaald zijn:	Items die in er in de tweede fase bij gekomen zijn:
atheïsme	bij elkaar op de koffie gaan (met een koekje)
gierig zijn	dichtbevolkt
koopmansgeest	discriminatie
orthodox protestantisme	files
seculier	gastvrijheid
Unie van Utrecht (1579)	handelsgeest
vrijzinnig protestantisme	het Rijksmuseum
Herman Willem Daendels	meedoen aan vredesmissies
Khadija Arib	niet godsdienstig zijn
boeken Hafid Bouazza	ontwikkelingshulp
	oranjegekte
	protestantisme
	traditioneel zijn
	zorgzaamheid
	zuinigheid

Bron: SCP, DAN'18

De uiteindelijke versie van de DAN-enquête is voorgelegd aan ruim 5500 respondenten van het LISS-panel. In deze enquête zijn 185 vragen opgenomen die verschillende kenmerken van Nederland bevatten. Zowel de vraag wat men typerend vindt voor Nederland, als wel de vraag wat men belangrijk vindt voor de verbondenheid met Nederland wordt gemeten op een vijfpuntsschaal,<sup>1</sup> waaraan ook de opties 'dit ken ik niet' en 'dit weet ik niet' toegevoegd zijn. Het volledige vragenblok is opgenomen in hoofdstuk 5. De antwoordcategorieën 'dit ken ik niet' en 'dit weet ik niet' kunnen moeilijk een plaats toegekend krijgen tussen de vijf andere categorieën. Middels een *Categorical Principal Component Analysis*<sup>2</sup> is onderzocht of we de schaal van de categorieën zouden kunnen aanpassen, opdat ook de categorieën 'dit ken ik niet' en/of 'dit weet ik niet' een plaats toegekend kunnen krijgen. Deze analyses zijn uitgevoerd op de typerend- en verbondenheid-items, waarna per item de optimale schalingsmethode is vastgesteld. De uitkomsten waren niet eenduidig genoeg om de antwoordschalen aan te passen. Zodoende is er in beide analyses voor gekozen om de antwoordcategorieën 'dit ken ik niet' en 'dit weet ik niet' buiten beschouwing te laten. Dit betekent dat antwoorden die in een van deze categorieën vallen als een ontbrekende waarneming beschouwd worden. Deze keuze heeft twee gevolgen.

Ten eerste is het mogelijk dat er een meetfout in onze analyses geïntroduceerd is, doordat we voorselecteren op een groep met mogelijk non-random kenmerken. Dit kan (maar hoeft geen) gevolgen te hebben voor de generaliseerbaarheid van de resultaten. We testen op het bestaan van een meetfout middels drie Tobit regressieanalyses, waarbij de afhankelijke variabele het aantal typerend- of verbondenheid-items telt, dat ofwel met 'ik weet het niet', ofwel met 'ik ken het niet' beantwoord is. Als voorbeeld nemen we het aantal items dat met 'ik weet het niet' beantwoord is. De afhankelijke variabele is gelijk aan 0 als de respondent alle 185 items op de vijfpuntsschaal beantwoord heeft, 1 als één van de vragen met 'ik weet het niet' beantwoord is, en loopt op naar 185 als alle items met 'ik weet het niet' beantwoord zijn. We voegen verschillende sociodemografische kenmerken toe om te kijken of deze de kans op een hoger aantal 'onbeantwoorde' items doet toenemen. We zien voor alle drie de regressieanalyses dat vrouwen vaker dan mannen een item met 'ik weet het niet' of 'ik ken het niet' beantwoorden. Dit verschil komt neer op twee à drie vragen. Ook zien we dat ouderen en hogeropgeleiden minder vaak vragen met 'ik weet het niet' beantwoorden. Van deze groep beantwoorden alleen de ouderen een vraag minder vaak met 'ik ken het niet'. Ook zien we een marginaal verschil tussen respondenten die zich tot een autochtoon Nederlandse groep rekenen en respondenten die zich tot een andere groep rekenen. De laatste groep heeft een marginale positieve kans om meer items met 'ik weet het niet' of 'ik ken het niet' te beantwoorden.

Ten tweede zien we dat het aantal ontbrekende waarnemingen per item soms enorm toeneemt als gevolg van onze keuze. Voor sommige items zou dit betekenen dat meer dan 40% van de observaties ontbreekt. Dit zou vooral voor de factoranalyses grote consequenties hebben, namelijk dat ongeveer de helft van het totaal aantal respondenten niet meegenomen kan worden in de analyses.<sup>3</sup> Dit beschouwen we als een te grote groep die afvalt. We hebben daarom de volgende twee stappen ondernomen.

Ten eerste hebben we items met veel ontbrekende waarnemingen uit de analyses gehaald. Een argument hiervoor is dat de items met veel ontbrekende waarnemingen weinig beantwoord zijn en dus weinig informatie bevatten. Als criterium houden we meer dan 750 ontbrekende waarnemingen op de verbondenheid-vraag aan. Dit betekent dat 21 vragen niet meegenomen zijn in de factoranalyses (164 items blijven over voor de analyses).<sup>4</sup> Het merendeel van deze vragen betreft historische gebeurtenissen en verbintenissen, zoals het Plakkaat van Verlatinghe, en historische figuren, zoals Kenau Simonsdochter Hasselaer. Tabel 2 geeft alle items weer, samen met het aantal ontbrekende waarnemingen per vraag. Deze tabel is ook opgenomen in hoofdstuk 18 en kort besproken in hoofdstuk 5. Het uitsluiten van deze items heeft weinig tot geen effect op de resultaten van de factoranalyses. Niet-getoonde resultaten laten zien dat we

---

<sup>1</sup> Voor de typerend-vraag: 1. Helemaal niet typerend – 5. Heel erg typerend. Voor de verbondenheid-vraag: 1. Draagt helemaal niet bij aan verbondenheid – 5. Draagt heel erg bij aan verbondenheid.

<sup>2</sup> Het commando *catpca* in SPSS.

<sup>3</sup> Alleen de respondenten die alle items ingevuld hebben, worden meegenomen in de correlatiematrix op basis waarvan de factoranalyse uitgevoerd wordt.

<sup>4</sup> Ook voor deze items hebben we gekeken of er een mogelijke meetfout bestaat. We vinden resultaten die overeenkomen met de analyses van de 185 items.

hetzelfde aantal dimensies vinden als we de 21 items met veel ontbrekende waarnemingen ook meenemen in de analyses. De dimensies komen ook inhoudelijk overeen met de oplossingen die in de tabellen A1 en A2 (zie appendix aan het einde van deze bijlage) zijn weergegeven.

Tabel 2

Items met meer dan 750 ontbrekende waarnemingen

	verbonden		typerend	
	absoluut	%	absoluut	%
censuur	751	15,4	510	10,4
Willem Drees	757	15,5	654	13,4
risicomijdend zijn	763	15,6	443	9,1
Srebrenica	815	16,7	666	13,6
gedegenheid	822	16,8	545	11,2
de Beeldenstorm en Reformatie	844	17,3	717	14,7
polderen	846	17,3	585	12,0
ironie	912	18,7	632	12,9
Desiderius Erasmus	916	18,8	808	16,5
Bonifatius	1004	20,6	864	17,7
Aletta Jacobs	1119	22,9	1043	21,4
Jan Pieterszoon Coen	1124	23,0	1032	21,1
Victory Boogie Woogie van Mondriaan	1127	23,1	1023	20,9
Spinoza	1147	23,5	1042	21,3
Brainport Eindhoven	1257	25,7	1122	23,0
de moord op de gebroeders De Witt	1270	26,0	1150	23,5
de Vrede van Münster (1648)	1341	27,5	1247	25,5
Robbert Dijkgraaf	1575	32,2	1507	30,8
de Bataafse opstand	1641	33,6	1540	31,5
Kenau Simonsdochter Hasselaer	2033	41,6	1994	40,8
het Plakkaat van Verlatinghe (1581)	2166	44,3	2153	44,1

Bron: SCP, DAN'18

Ten tweede zijn alle factoranalyses geschat met *full information maximum likelihood* (FIML). Deze techniek is beschikbaar in Mplus en berekent aannemelijkheidsfuncties voor elke respondent, onafhankelijk van de ontbrekende waarnemingen (Enders 2001). Een dergelijke aannemelijkheidsfunctie berekent de kans op een bepaalde uitkomst gegeven de beschikbare informatie. Zo worden ook de respondenten met een ontbrekende waarneming meegenomen<sup>5</sup> in de berekeningen, opdat aan elke respondent een uiteindelijke factorwaarde toegekend kan worden. Deze techniek maakt geen gebruik van imputatie, noch van *pairwise deletion* en wordt gezien als superieur aan deze technieken, omdat het zich baseert op alle beschikbare informatie.

### 1.2 Methode

In tabel 3 vatten we de stappen samen die we hebben gevolgd bij onze analyses. In een eerste stap wordt zowel exploratieve (EFA) als bevestigende (ook wel confirmatieve) factoranalyse (CFA) gebruikt om de dimensionaliteit van de Nederlandse identiteit vast te stellen en te valideren (Wansbeek en Meijer 2000). We volgen hiervoor een tweeledige opzet. Allereerst gebruiken we EFA om op exploratieve wijze de

<sup>5</sup> Zie tabel B1 in de bijlage van hoofdstuk 5 voor een overzicht van alle items en waarnemingen. Waarnemingen zijn geteld als een antwoord is gegeven op de vijfpuntsschaal. Ontbrekende waarnemingen kunnen geteld worden door N af te trekken van het totaal aantal observaties (4885).

dimensionaliteit van identiteit vast te stellen. Dit wordt gedaan zonder a priori restricties op te leggen aan de hoeveelheid factoren. Enerzijds kijken we naar wat Nederlanders typerend vinden voor Nederland, anderzijds kijken we naar wat Nederlanders verbindt met Nederland. Om de robuustheid van deze oplossing vast te stellen gebruiken we een confirmatieve factoranalyse. Deze analyse valideert het EFA-model. Volgend op de factoranalyses van de verbondenheid-items gebruiken we de archetypenanalyse (AA) om archetypen, oftewel een typologie, van Nederlandse identificatie te maken in hoofdstuk 8.

De methode is in hoge mate verbonden met de opzet van de dataverzameling. Zoals beschreven is er voorafgaand aan de steekproef die is uitgezet binnen het LISS-panel een enquête uitgezet via het I&O-panel. In deze testronde is een versie van de DAN-enquête getest die later is aangepast en verder verfijnd. Omdat de steekproef van het LISS-panel te verkiezen is boven die van het I&O-panel en er een discrepantie bestaat tussen de twee vragenlijsten, kiezen we ervoor om de steekproef van het LISS te gebruiken voor onze analyses. We splitsen het databestand in tweeën via een random gegenereerde getallenreeks. We gebruiken 50% van de respondenten voor de exploratieve analyses en de resterende 50% voor de confirmatieve analyses. De data die verkregen zijn vanuit het I&O-panel gebruiken we om de robuustheid van onze factoranalyses vast te stellen. Dit geeft ons de volgende opzet:

Tabel 3

Overzicht methodologische stappen

Stap 1	exploratieve analyse van een at random geselecteerde 50% van het LISS-panel
Stap 2	exploratieve analyse van het I&O-panel als robuustheidstest van stap 1
Stap 3	keuze voor exploratief factormodel gebaseerd op de uitkomsten LISS
Stap 4	testen van het exploratieve model middels confirmatieve factoranalyse van de resterende 50% van het LISS-panel
Stap 5	verfijning model op basis van het confirmatieve model
Stap 6	schatten van het finale model op het gehele LISS-panel

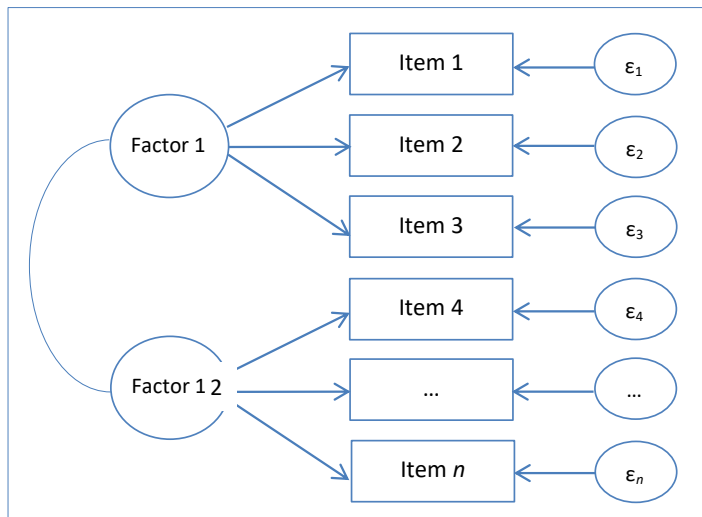
Hierna beschrijven we de methodologie in meer detail, voordat we de resultaten van de verschillende stappen uiteenzetten.

#### *Exploratieve factoranalyse*

Om vast te kunnen stellen of de waargenomen variabelen – de typerend- en verbondenheid-items – afhangen van onderliggende latente variabelen wordt de correlatie tussen de items onderzocht (Bartholomew, Steele, Galbraith en Moustaki 2008). In de context van het SCR wordt verondersteld dat de latente factoren enerzijds identiteit (typerend-items) en anderzijds identificatie (verbondenheid-items) behelzen. Figuur 1 geeft een illustratie van een mogelijke factorstructuur. In deze structuur worden de items door twee factoren verklaard. Figuur 1 laat zien dat de items 1 tot en met 3 verklaard worden door factor 1. Dit betekent dat ze op een bepaalde manier samenhangen en dat hun samenhang verschilt van die van items 4, ...,  $n$ . Deze items worden namelijk verklaard door factor 2. Het is mogelijk dat de twee factoren met elkaar samenhangen (een zekere mate van correlatie). Zeker als het gaat om sociale fenomenen, komt dit vaak beter overeen met de werkelijkheid dan factoren die niet samenhangen. Zoals eerder gezegd, verklaren de factoren dus de variantie die items gemeenschappelijk hebben. De variantie die overblijft, wordt aangegeven door verstoringstermen ( $\epsilon_n$  in figuur 1) en kan gezien worden als de variantie die uniek is voor dat item.

Figuur 1

Weergave van een mogelijke factorstructuur



We kunnen het factormodel ook analytisch weergeven. Het factormodel voor meerdere factoren  $k$  en indicatoren  $m$  ziet er als volgt uit:

$$\begin{aligned}
 Y_i &= B\xi_i + \epsilon_i \\
 \xi_i &\sim N(0, \Phi) \\
 \epsilon_i &\sim N_m(0, \Omega)
 \end{aligned}$$

$Y_i$  en  $\epsilon_i$  zijn kolomvectoren met  $m$  elementen en  $\xi_i$  is een kolomvector met  $k$  elementen.  $B$  is een matrix van  $m \times k$ . Er wordt aangenomen dat  $\xi_i$  en  $\epsilon_i$  onafhankelijk en multivariaat normaal verdeeld zijn,  $\Phi$  is een positief-definiete matrix van  $k \times k$  en  $\Omega$  is een positief-definiete matrix van  $m \times m$ . Hieruit volgt dat  $Y_i \sim N_m(0, \Sigma)$  met  $\Sigma \equiv B\Phi B' + \Omega$ . Nu hebben we dus een factoroplossing met factorwaarden in matrix  $B$  voor  $k$  factoren,  $\xi_i$  een vector met factorscores en een vector  $Y_i$  waarin de indicatoren weergegeven zijn. De factorwaarden representeren de samenhang tussen de indicator en de latente factor. Daarom zijn alleen factorwaarden die hoog genoeg zijn informatief; het zijn deze indicatoren die worden meegenomen in de uiteindelijke interpretatie van de factoren.<sup>6</sup>

We voeren de EFA uit op de 164 geselecteerde enquête-items via Mplus. Wij houden een schatting met *maximum likelihood* aan. We roteren de factoroplossing (Bartholomew et al. 2008). We laten toe dat factoren correleren, omdat het zeer aannemelijk is dat de verschillende dimensies van identiteit met elkaar samenhangen.

Normaliter gebruikt men verschillende selectiecriteria om het aantal factoren vast te kunnen stellen. Zo wordt de *Cattell's scree test* gebruikt, die op basis van een grafische weergave van het aantal factoren ten opzichte van de eigenwaarden een beste oplossing aanduidt (Cattell 1966). De knik in de grafiek bepaalt het aantal factoren dat behouden blijft. We kunnen ook het *Kaiser-criterium* gebruiken, dat zich baseert op de eigenwaarden. Alle factoren met een eigenwaarde hoger dan één blijven dan behouden. Ook kan er gekeken worden naar de *Bayesian Information Criterion* (BIC) en de *Root Mean Squared Error of Approximation* (RMSEA). Een lagere BIC duidt een beter model aan en de RMSEA zou onder de 0,05 moeten liggen (Preacher et al. 2013). Een waarde groter dan 0,10 duidt op een 'slecht' model. Ook een *chi-squared test* kan uitwijzen hoeveel factoren behouden moeten blijven. Deze test gaat uit van de nulhypothese dat het geschatte model een goede representatie van de populatie is. Afwijzing van de nulhypothese betekent dat een model met een

<sup>6</sup> Een factorlading met een absolute waarde van minimaal 0,4 is als vuistregel aangehouden.



extra factor geschat moet worden. De test is gevoelig voor de steekproefomvang en kan bij een groot aantal observaties niet langer betrouwbaar zijn. Om deze reden wordt dit criterium niet gebruikt voor de selectie van het optimale model.

Naast selectiecriteria is ook de betekenis van het model belangrijk. Het kan namelijk goed zijn dat factoren slecht te interpreteren zijn zodra een bepaalde drempel overschreden wordt. Uiteindelijk moeten factoren geïnterpreteerd kunnen worden binnen het kader van het voorgeschreven onderzoek.

#### Confirmatieve factoranalyse

Daar waar EFA vertrekt zonder a priori hypothesevorming, vormen hypothesen de basis voor een bevestigende factoranalyse (CFA). CFA wordt zodoende vaak ingezet om hypothesen die gebaseerd zijn op theorie(en) te testen en te valideren. Dit betekent dat voor de onderzoeker met de analyses begint, zij duidelijke ideeën heeft over het aantal factoren, hun onderlinge relatie en welke indicatoren veroorzaakt worden door welke latente factor. Een voorbeeld van een dergelijk model staat in figuur 1, ervan uitgaande dat die gebaseerd is op een theorie. Naast een theorie kan ook de uitkomst van een EFA gebruikt worden om de restricties te bepalen. Zo worden EFA en CFA vaak in combinatie gebruikt, opdat CFA de structuur van EFA kan valideren. Ook in dit onderzoek hebben we deze methode gevolgd.

Het analytische model is hetzelfde als dat van EFA. Maar, zoals eerder benoemd, worden in het geval van CFA restricties gelegd op matrix **B**. In het SCR hebben we de onderlinge relaties tussen factoren en indicatoren gebruikt om het bevestigende model te kunnen schatten. Dit betekent dat items alleen op één factor kunnen laden en niet, zoals in het geval van EFA, op verschillende factoren. Evenals met de schattingen van de EFAs hebben we ook hier van Mplus gebruikgemaakt (*maximum likelihood*). Meer specifiek, we schatten een Structural Equation Model (SEM), dat het EFA-model test.

Ook voor een CFA bestaan er verschillende criteria die gebruikt worden om het model te evalueren. Allereerst kan een *Likelihood Ratio (LR) test* gebruikt worden die de opgelegde restricties test ten opzichte van het model zonder restricties. De nulhypothese stelt dat het opgelegde model goed past. Ten tweede kan een *Comparative Fit Index (CFI)* aanduiden welk model relatief gezien het beste past. Over het algemeen wordt aangenomen dat een CFI hoger moet zijn dan 0,90 en dat relatief gezien het model met de hoogste CFI het best passend is. In de derde plaats is ook de *BIC* een comparatieve maatstaf. Voor modellen die dezelfde variabelen bevatten, wordt het model met de laagste BIC als het beste beschouwd. Ten vierde wordt ook voor CFA de *RMSEA* gebruikt, waarbij een waarde kleiner dan 0,05 een passend model aanduidt. Als laatste gebruiken we de *Standardized Root Mean Squared Error (SRMSE)*. Deze index, die de covariantie van de residuen gebruikt, is idealiter lager dan 0,08 voor een goed passend model (Hu en Bentler 1999).

### 1.3 Resultaten

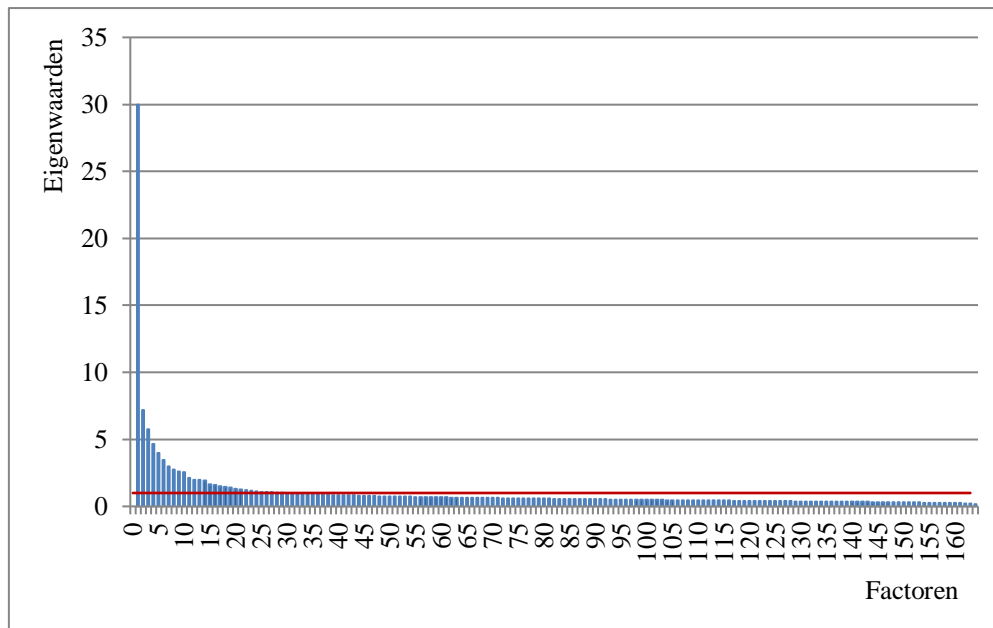
#### 1.3.1 Typerend voor Nederland

##### Exploratieve fase

Op basis van het LISS-bestand selecteren we 2442 observaties op willekeurige wijze. Dit bestand vormt de basis voor onze exploratieve analyses van de typerend-items. Om het optimale factormodel te schatten, kijken we allereerst naar de *scree plot* van de eigenwaarden. Figuur 2 laat een dergelijke plot zien, waarbij de eigenwaarden (verticale as) afgezet worden tegen het aantal factoren (horizontale as). Cattell's scree test schrijft voor dat de 'knik' in de grafiek het optimale aantal factoren aanduidt. Er is geen duidelijke knik te zien, maar rond de twaalf/dertien factoren lijkt er een flauwe afbuiging te bestaan. De figuur geeft ook aan dat na 29 factoren de eigenwaarden lager dan één worden (aangeduid door middel van de rode lijn in figuur 2), zoals het Kaiser-criterium voorschrijft.

Figuur 2

Plot van de eigenwaarden ten opzichte van het aantal factoren (LISS)



Bron: SCP, DAN'18

Aangezien de verschillende fitmaten niet eenduidig zijn over het optimale aantal factoren, schatten we een reeks exploratieve factormodellen op de helft van het LISS-bestand en op het gehele I&O-bestand (als robuustheidstest). Dit opdat we naast de andere selectiecriteria (BIC, RSMSEA, CFI, TLI, *chi-squared test*) ook de interpretatie van de verschillende modellen naast elkaar kunnen leggen. We zijn immers op zoek naar een meting van Nederlandse identiteit die naast statistisch valide ook interpretabel is.

Eerst de analyses van het LISS-bestand. We schatten verschillende factormodellen met vijf tot en met twintig factoren. We roteren de oplossingen zo, dat er correlatie tussen de factoren toegestaan is (*geomin rotatie*). De resultaten zijn beschikbaar in een online appendix. Tabel A1 in de appendix geeft de factormodellen weer waarop het model gepresenteerd in hoofdstuk 7 gebaseerd is.

We analyseren en interpreteren de verschillende modellen. Wederom geven de selectiecriteria geen eenduidig beeld. Kijkend naar de RMSEA, zien we dat deze voor alle modellen onder de 0,05 valt. Oftewel, elk model is valide. De BIC blijft dalen naarmate er meer factoren toegevoegd worden, wat aangeeft dat het model met één extra factor steeds geprefereerd wordt. De *chi-squared test* geeft eenzelfde beeld. Bij elk model wordt de nulhypothese verworpen; ook deze test geeft de voorkeur aan een model met meerdere factoren. De CFI en TLI zouden boven de 0,90 moeten reiken om een goed model aan te duiden. In geen van de modellen is dit het geval. Mogelijk kan dit verklaard worden doordat het nulmodel al redelijk goed is. We komen hier later op terug wanneer we de CFA bespreken.

De resultaten laten wel zien dat na vijftien factoren de factoroplossingen minder interpretabel worden. Zo ontstaan er bijvoorbeeld factoren die op weinig items laden of lastig te interpreteren zijn. We kiezen er daarom voor om het model met vijftien factoren als uitgangspunt te nemen voor de confirmatieve analyses. De verschillende factoren van het vijftien-factorenmodel geven een duidelijk en goed te interpreteren beeld van de onderliggende factorstructuur.

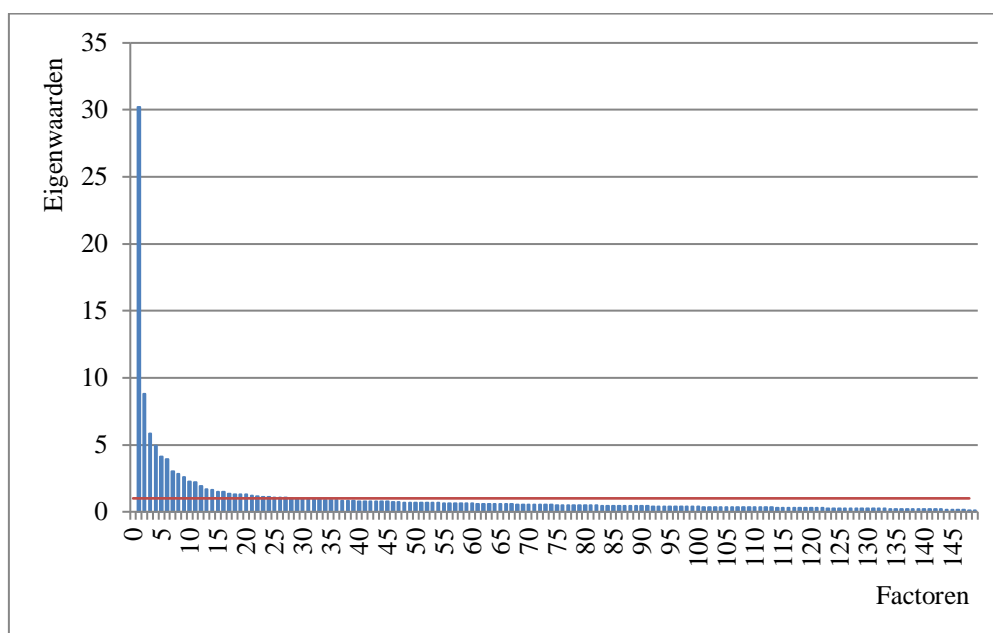
Naast de analyses van het LISS-bestand voeren we eenzelfde analyse uit op het I&O-bestand. Gezien de grootte van de steekproef van het LISS en de representativiteit daarvan, zien we de analyses van het I&O-bestand als ondersteunend en bespreken we de uitkomsten minder uitvoerig. We volgen dezelfde opzet als die bij de analyses van het LISS-bestand. De input voor de analyses zijn de 149 items die het I&O- en LISS-bestand

delen nadat de 21 items met de meeste missings verwijderd zijn, zoals besproken in paragraaf 1.1 van deze bijlage.

Figuur 3 plot wederom de eigenwaarden ten opzichte van het aantal factoren. De structuur van de eigenwaarden komt overeen met die van de analyse van het LISS-bestand. Net als daar zien we ook hier geen duidelijke knik in de figuur, wat ons zou kunnen helpen het optimale factormodel te bepalen. Wel zien we een flauwe buiging rondom twaalf/dertien factoren. Na 28 factoren worden de eigenwaarden kleiner dan 1. Andere selectiecriteria (BIC, RSMSEA, CFI, TLI, chi-squared test) schetsen eenzelfde beeld als bij de analyses van het LISS-panel. Wat we ook terug zien komen, is dat na ongeveer vijftien factoren de modellen minder goed te interpreteren zijn. Zo ontstaan ook hier factoren die lage ladingen hebben of zijn er items die samen een groep vormen terwijl ze inhoudelijk weinig met elkaar te maken hebben.

Figuur 3

Plot van de eigenwaarden ten opzichte van het aantal factoren (I&O)



Bron: SCP, DAN'18

Als we de oplossingen van het I&O- en LISS-panel met elkaar vergelijken, dan zien we dat er, ondanks de verschillen in volgorde waarin de items aan de respondenten zijn voorgelegd, vergelijkbare factoren gevonden worden. Zo komen de meeste factoren, zoals *democratie*, *vrijheden en rechtstaat*, *tradities en gewoonten*, *kunst*, *geschiedenis* en *beroemdheden*, terug onafhankelijk van het databestand. Dit onderschrijft de stabiliteit van de bevindingen. Er bestaan echter ook verschillen tussen de oplossingen die gevonden worden tussen de databestanden. Kijken we naar de modellen met vijftien factoren, dan zien we in de analyses van het LISS-bestand een factor die gerelateerd wordt aan directheid in communicatie, botheid, niet godsdienstig zijn, het gedoogbeleid en coffeeshops; een factor die de Nederlandse liberale houding lijkt te vatten. Deze factor komt niet terug in de analyses van het I&O-bestand, net als de factor die gerelateerd is aan de *verzorgingsstaat*. In de analyses van het I&O-bestand komt er een factor gerelateerd aan Kerstmis, de Koude Oorlog en het jodendom naar voren; een factor die moeilijk te duiden is. Ook zien we dat de sportprestaties gescheiden worden van het gevoel en de emotie, zoals gekenmerkt door oranjegekte en oranjegevoel. Dit zien we niet terug in de analyses van het LISS-bestand. Mogelijk kunnen de verschillen verklaard worden door de verschillen tussen de twee steekproeven, maar dit is lastig om te testen.

Doordat de factoroplossingen ook in het geval van het I&O-bestand diffuser worden naarmate we modellen met meer dan vijftien factoren schatten, zien we dit als onderschrijving van het vijftien-

factorenmodel dat uit de analyses van het LISS-bestand komt. We nemen daarom dit model als uitgangspunt voor de confirmatieve analyse.

#### Confirmatieve fase

We schatten een confirmatief factormodel gebaseerd op het vijftien-factorenmodel dat we op de helft van het LISS-bestand geschat hebben. De confirmatieve analyse zal uitwijzen of deze items daadwerkelijk verklaard worden door deze onderliggende constructen. Items die te lage ladingen hebben ( $< |0,4|$ ) zullen op iteratieve wijze afvallen om zodoende het model te verfijnen. De lage ladingen signaleren dat deze items mogelijk niet door de onderliggende factor verklaard worden en kunnen daarom buiten beschouwing blijven.

We gebruiken de 2442 willekeurig geselecteerde observaties uit het LISS-bestand als input voor de analyses. Wederom schatten we het model in Mplus, opdat we ook voor de ontbrekende waarnemingen een factorscore kunnen schatten. De oplossing wordt geroteerd middels een non-orthogonale rotatie, wat betekent dat correlatie tussen de verschillende factoren is toegestaan.

Het geschatte vijftien-factorenmodel bevestigt grotendeels de gevonden exploratieve factorstructuur. We vinden stabiele factoren voor *vrijheden*, *democratie en rechtstaat*, *deugden*, *islamitische traditie*, *sportprestaties*, *de verzorgingsstaat*, *symbolen*, *beroemdheden*, *landschap* en *kunst*. Voor de factoren *ondeugden*, *directheid* en *waterbeheer* zijn er verschillende items met lage ladingen. Vooral de factor *directheid*, die ook al lage ladingen had in het EFA-model, blijkt instabiel te zijn. We schatten daarom een nieuw model zonder deze factor; dit keer dus een CFA met veertien factoren.

We verfijnen het veertien-factorenmodel, dus *zonder* de factor *directheid*, op basis van de factorladingen en fitmaten. De items met te lage ladingen ( $< |0,4|$ ) laten we stapsgewijs afvallen. Dit betekent dat de volgende items afvallen: corruptie, achterkamertjespolitiek, internationale oriëntatie, eten prakken, en Ayaan Hirsi Ali. Deze items hadden in de eerdere specificaties (EFA vijftien-factorenmodel en CFA vijftien-factorenmodel) ladingen die rond het criterium van 0,4 lagen. Dit geeft aan dat ook in deze modellen onduidelijk is of de items wel of niet door de onderliggende factoren verklaard konden worden. Dat de lading uiteindelijk onder de  $|0,4|$  valt, is daarom doorslaggevend.

Het optimale veertien-factorenmodel bestaat dus uit de volgende veertien, stabiele factoren: *vrijheden*, *ondeugden*, *deugden*, *tradities*, *islamitische traditie*, *sportprestaties*, *verzorgingsstaat*, *joods-christelijke traditie*, *geschiedenis*, *symbolen*, *beroemdheden*, *landschap*, *waterbeheer* en *kunst*.<sup>7</sup> Dit model is geschat op 126 items. Alle factorladingen zijn in dit model hoog genoeg ( $> |0,4|$ ), waarvan de meerderheid ruimschoots boven dit criterium uitkomt (90% van de items heeft een lading  $> |0,50|$ ). De RMSEA is laag genoeg ( $< 0,05$ ), SRMR valt onder de 0,8 (Asparouhov en Muthén 2009; Asparouhov en Muthén 2018), alleen de CFI en TLI blijven relatief laag (0,76 en 0,75).<sup>8</sup> Beide waarden zouden boven de 0,90 uit moeten komen, maar vallen hier ruimschoots onder. Een mogelijke verklaring ligt in de eigenschappen van het nulmodel, waarmee het CFA-model vergeleken wordt (Kenny 2014). Als het nulmodel – een model waarbij alle paden tussen items en factoren gelijkgesteld worden aan 1, en de varianties van en correlaties tussen factoren gelijkgesteld worden aan 0 – al behoorlijk goed past, dan lijden de CFI en TLI daaronder. Beide maatstaven duiden namelijk de verbetering ten opzichte van het nulmodel. Deze verbetering is gering als het nulmodel al goed passend is. Schatting van het nulmodel laat zien dat de RMSEA met een waarde van 0,071 inderdaad veel lager is dan 0,158. Dit verklaart dus waarom de CFI en TLI niet boven de 0,90 kunnen komen.

We schatten het optimale veertien-factorenmodel als een finale test op het gehele LISS-panel, zodat we deze uitkomsten kunnen gebruiken om de factorscores voor alle respondenten van de DAN-enquête te schatten. Als alle ladingen inderdaad hoog genoeg blijven ( $> |0,4|$ ) en de fitmaten goed genoeg zijn, dan bevestigt dit wederom ons model. Dit blijkt inderdaad het geval te zijn, zoals de fitmaten in tabel A1 (zie de appendix einde van deze bijlage) illustreren. Ook duiden de Cronbach's alfa's aan dat de dimensies intern

---

<sup>7</sup> We hebben de stabiliteit van dit model ook getest, gebruikmakend van een zogenoemde ESEM (Exploratory Structural Equation Model (Asparouhov en Muthén 2009; Marsh et al. 2010; Marsh, Morin, Parker en Kaur 2014; Marsh, Muthén, Asparouhov, Lüdtke en Robitzsch 2009; Morin, Marsh en Nagengast 2013). De resultaten zijn vergelijkbaar en kunnen zodoende als een robuustheidstest van ons confirmatieve model gezien worden.

<sup>8</sup> Hetzelfde geldt voor het model met vijftien factoren en voor het veertien-factorenmodel dat geschat is op alle geselecteerde items.

consistent zijn (nabij of hoger dan 0,80) en nagenoeg dezelfde onderliggende dimensie meten. Alleen de factor ondeugden heeft een alfa die onder de 0,80 valt. Met een waarde van 0,64 stellen we dat de interne consistentie voor deze factor matig is. Voor de analyses in de hoofdstukken 7 en 8 van dit SCR, die gebruikmaken van dit factormodel, schatten we voor iedere dimensie per respondent de gemiddelde score van de items die laden op de dimensie. Respondenten die minder dan de helft van het aantal items per dimensie beantwoord hebben, worden als een ontbrekende waarneming meegenomen. Dit geldt voor 15% van het gehele sample. De respondenten die hierdoor uitvallen, lijken niet door demografische of politieke variabelen (stemgedrag) beïnvloed te zijn. De verkregen gemiddeldes sluiten aan op de vijfpuntsschaal uit de enquête. In het SCR wordt daarom alleen naar deze gemiddelde scores verwezen en niet naar de factorladingen of factorscores.

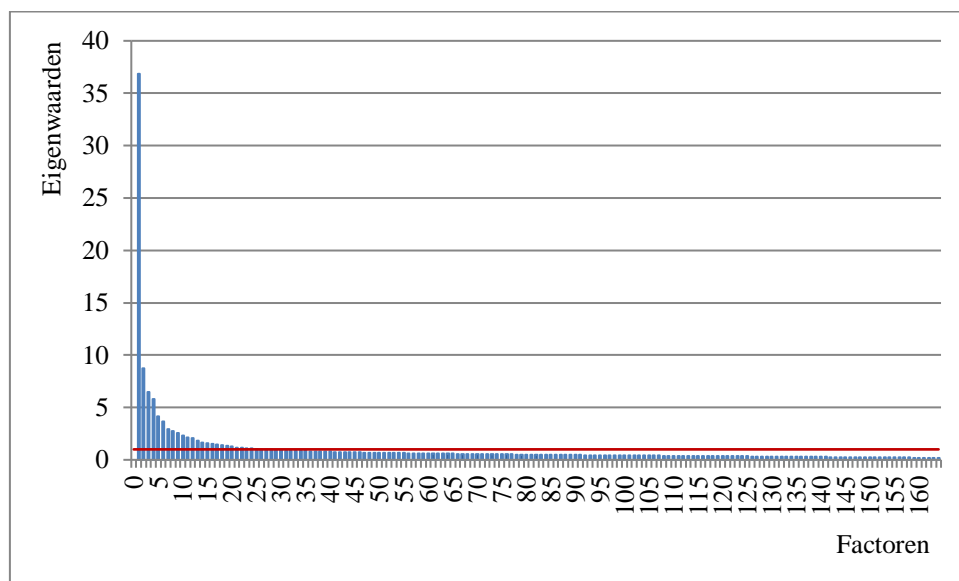
### 1.3.2 Verbondenheid met de Nederlandse identiteit

#### Exploratieve fase

We volgen een opzet gelijk aan die van de analyses van de typerend-items. Dit betekent dat we een reeks exploratieve analyses uitvoeren op dezelfde helft van het LISS-bestand zoals eerder gebruikt (2442 observaties). Dezelfde analyses voeren we uit op het I&O-panel. Deze worden gebruikt ter ondersteuning van de analyses van het LISS-panel om eerdergenoemde redenen.

Figuur 4

Plot van de eigenwaarden ten opzichte van het aantal factoren (LISS)



Bron: SCP, DAN'18

Om het optimale factormodel te schatten kijken we wederom naar de *scree plot* van de eigenwaarden. Figuur 4 laat deze plot zien. De figuur laat geen duidelijke knik zien, maar een afbuiging van de waarden lijkt rond de tien factoren te bestaan. De tweede factor reduceert de variantie in de data duidelijk; we zien een afname van bijna 30 punten. Dit komt overeen met patronen die we in de eerdere *scree plots* zagen (figuur 3 en 4). De rode lijn duidt eigenwaarden aan met een waarde lager dan 1, het standaard ijkpunt voor het bepalen van het aantal factoren. Bij meer dan 25 factoren worden de eigenwaarden lager dan 1.

Wederom geven Cattell's scree test en het Kaiser-criterium beide geen eenduidig antwoord. We schatten daarom een reeks exploratieve factormodellen (vijf tot twintig factoren). Dit opdat we naast de andere selectiecriteria (BIC, RSMSEA, chi-squared test) ook de interpretatie van de verschillende modellen

naast elkaar kunnen leggen. De oplossingen worden niet-orthogonaal geroteerd, opdat de factoren met elkaar kunnen correleren.

Laten we eerst de analyses van het LISS-bestand verder uitdiepen. De fitmaten zijn gunstiger voor een model met meerdere factoren. We zien dat naarmate het aantal factoren toeneemt, de RMSEA en SRMR lager worden en de CFI en TLI toenemen. Voor alle modellen zijn de waarden van de RMSEA en SRMR echter laag genoeg, dat wil zeggen kleiner dan 0,05 respectievelijk 0,08. Dit betekent dat, als we alleen naar deze twee fitmaten zouden kijken, alle modellen als valide gezien zouden kunnen worden. De CFI en TLI zijn vaak net te laag, met waarden die steeds dichterbij de 0,90 komen naarmate het aantal factoren toeneemt. Wederom zou dit kunnen, omdat het nulmodel te goed is. We komen hierop terug bij de confirmatieve analyses.

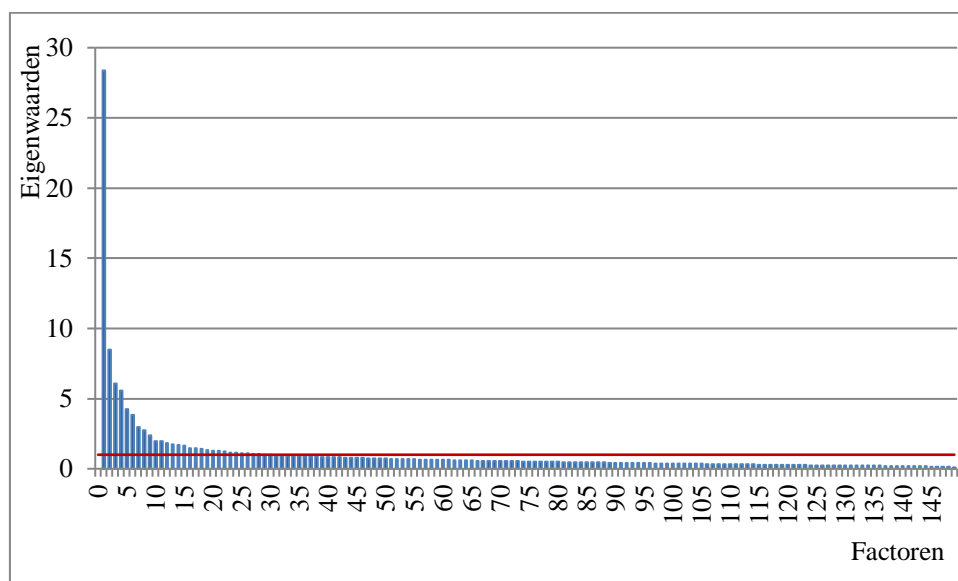
We zien dat de modellen met vijftien en meer factoren minder goed te interpreteren zijn dan de modellen met dertien of veertien factoren. Er ontstaan factoren die op weinig items laden en lage ladingen toegewezen krijgen. Dit duidt erop dat deze modellen minder stabiel zijn dan die met minder factoren.

Gebaseerd op de fitmaten en de inhoud van de verschillende modellen (en omdat modellen met minimaal vijftien factoren slecht zijn te interpreteren) houden we het vijftien-factorenmodel, samen met het veertien-factorenmodel, aan als uitgangspunt voor de confirmatieve analyses. Voor we dit doen, kijken we echter of de analyses van het I&O-bestand eenzelfde beeld schetsen. Als input gebruiken we de 149 items die het I&O- en LISS-bestand delen, nadat we de 21 items met de meeste ontbrekende waarnemingen hebben verwijderd. Dit is dezelfde set items als gebruikt voor de analyses van de typerend-items.

Figuur 5 plot de eigenwaarden ten opzichte van het aantal factoren. Deze figuur bevestigt het beeld dat figuur 4 al schetste, namelijk dat er geen duidelijke knik te zien valt, maar wel een afbuiging van waarden rond tien factoren. Ook de flinke afname van de variantie van de eerste naar de tweede factor kunnen we aflezen uit de figuur. Na 29 factoren vallen de eigenwaarden onder het ijkpunt van 1. Het is een bekend beeld dat figuur 5 schetst. Net als bij de andere gevallen schatten we daarom ook hier weer een reeks factormodellen (vijf tot en met twintig factoren), zodat we het optimale model kunnen bepalen en kunnen vergelijken met de oplossingen die we eerder vonden.

Figuur 5

Plot van de eigenwaarden ten opzichte van het aantal factoren (I&O)



Bron: SCP, DAN'18

De analyses van het I&O-panel volgen globaal gezien de structuur die we vastgesteld hebben bij het LISS-panel. Dit betekent dat we factoren vinden die vergelijkbaar zijn. De factoren *vrijheden, democratie en rechtstaat, ondeugden, deugden, tradities en gewoonten, sportprestaties, islamitische traditie, joods-christelijke traditie, landschap, beroemdheden, geschiedenis, economie en kunst* zien we in de analyses van beide databestanden terug. De samenstelling van de factoren verschilt echter. Zo zien we bijvoorbeeld dat de factor *deugden* zich in twee factoren opsplijt wanneer we het LISS-bestand analyseren, terwijl deze als één factor geschat wordt in de analyses van het I&O-bestand. Ook de factor *tradities en gewoonten* wordt op verschillende manieren geschat. Het model dat gebaseerd is op het LISS-bestand presenteert een factor die te maken heeft met gewoonten en tradities, zoals sinterklaas en dodenherdenking. De analyses van het I&O-panel schatten deze factor veel minder breed, maar voegen wel een aantal aan symboliek gerelateerde items toe, zoals het koningshuis en de Nederlandse vlag. Hierdoor krijgt de factor een andere betekenis, namelijk een die veel meer gerelateerd is aan de Nederlandse monarchie en veel minder aan tradities en gewoonten in de brede zin van het woord. De factor *geschiedenis*, die als één factor in de analyses van het LISS-panel geschat wordt, valt in twee factoren uiteen in de analyses van het I&O-panel. Daar zien we dat er een factor ontstaat die verbonden is aan het handelsverleden van Nederland: de VOC, de Gouden Eeuw, Piet Hein en Michiel de Ruyter. De factor *symbolen* zien we helemaal niet meer terug in de analyses van het I&O-bestand. Deze items worden toegevoegd aan de factoren *landschap en tradities*.

Ondanks de verschillen tussen de modellen, overlapt de globale interpretatie van de modellen in grote mate. We concluderen daarom dat de modellen die geschat zijn op het LISS-panel robuust zijn en nemen de modellen met vijftien en veertien factoren als startpunt voor de confirmatieve analyses. Tabel A2 (opgenomen in de appendix aan het einde van deze bijlage) geeft een overzicht van de factorladingen van beide modellen.

#### Confirmatieve fase

Om het exploratieve model te testen, schatten we een confirmatief factormodel dat gebaseerd is op de factoruitkomsten van de modellen met veertien en vijftien factoren. Dit doen we, net als voor de modellen op de typerend-vraag, op de andere helft van het LISS-panel. Dit betekent dat we 2442 observaties gebruiken als input voor de analyse. De oplossingen worden geroteerd door een non-orthogonale rotatie. Items met factorladingen hoger dan  $|0,4|$  beschouwen we als informatief en behouden we in het model. Andere items vallen af. De modellen worden wederom in Mplus geschat, opdat ook de ontbrekende waarnemingen meegenomen worden in de analyses.

Het geschatte vijftien-factorenmodel bevestigt de gevonden exploratieve factorstructuur. Dertien van de vijftien factoren komen als stabiele factoren naar voren. Het gaat hier om *vrijheden, democratie en rechtstaat, ondeugden, deugden, tradities en gewoonten, islamitische traditie, joods-christelijke traditie, unieke sportprestaties, symbolen, beroemdheden, geschiedenis, landschap, economie en kunst*. Alleen de factoren die we in tabel A2 (opgenomen in de appendix aan het einde van deze bijlage) niet benoemd hebben, gelabeld als de factoren Y en Z, krijgen lage ladingen toegekend. Dit is niet verrassend, gezien hun inhoud; de items zijn namelijk ook inhoudelijk niet sterk met elkaar verbonden. We schatten een nieuw model zonder deze twee factoren, maar met een aparte factor voor items die gerelateerd zijn aan de verzorgingsstaat.<sup>9</sup> We nemen alle items mee die in het veertien- en vijftien-factorenmodel naar voren komen, omdat deze items inhoudelijk met elkaar verbonden zijn. De confirmatieve analyse zal laten zien of dit ook op statistische gronden hardgemaakt kan worden.

Net als bij het model op de typerend-vraag laten we items met een lading lager dan  $|0,4|$  stapsgewijs afvallen. De volgende items vallen af: corruptie, discriminatie, internationale oriëntatie, de NSB. Het optimale factormodel op de verbondenheid-vraag bestaat uit veertien factoren: *vrijheden en democratie, ondeugden, deugden, tradities en gewoonten, islamitische traditie, sportprestaties, verzorgingsstaat, joods-christelijke*

---

<sup>9</sup> De dimensie verzorgingsstaat kwam niet uit de EFA of CFA naar voren, maar uit een tussenanalyse (ESEM). Dit voert te ver om hier uiteen te zetten, maar omdat de twee modellen op deze manier inderdaad vergelijkbaar zijn en de factor stabiel is in de CFA-analyse, is deze dimensie toegevoegd aan het model (zie ook voetnoot 5).

*traditie, geschiedenis, symbolen, beroemdheden, landschap, economie en kunst.*<sup>10</sup> Dit model is gebaseerd op 132 items. Om factorscores voor alle respondenten in onze dataset te verkrijgen schatten we het model als een finale test op het gehele LISS-panel. De factorladingen van dit model zijn opgenomen in tabel A2 in de appendix aan het einde van deze bijlage. De tabel laat zien dat van de 131 items, 121 items een factorlading hebben die hoger is dan |0,5|. Tien items hebben een factorlading tussen de |0,4| en |0,5|. Ook voor dit model duiden de Cronbach's alfa's een hoge mate van interne consistentie aan. Nagenoeg alle alfa's zijn ruimschoots hoger dan de ijkwaarde van 0,80.

Voor het finale CFA-model zien we dat de RMSEA laag genoeg is met een waarde van 0,041. De CFI en TLI zijn, net als bij het model op typerend, te laag met een respectievelijke waarde van 0,738 en 0,732. Een mogelijke verklaring hiervoor zou gevonden kunnen worden in het schatten van het nulmodel. Als dit model 'te goed' is, dan lijden de CFI en TLI hieronder. Een schatting van het nulmodel laat zien dat het nulmodel inderdaad 'te goed' is. De kwaliteit van het model wordt vastgesteld middels de grootte van de RMSEA, die niet kleiner dan 0,158 moet zijn. Schatting van het nulmodel laat zien dat de RMSEA met een waarde van 0,078 inderdaad lager is dan 0,158. Dit verklaart dus waarom de CFI en TLI niet boven de 0,90 kunnen komen.

Voor de analyses in het SCR worden ook voor het model op verbondenheid gemiddelde scores per dimensie geschat. Hierdoor kan de interpretatie van de oorspronkelijke antwoordschalen aangehouden worden. Voor elke respondent die meer dan de helft van de items per dimensie beantwoord heeft, berekenen we een gemiddelde score. Respondenten die minder dan de helft van het aantal items per dimensie beantwoord hebben, worden als een ontbrekende waarneming meegenomen. Dit geldt voor 18% van de gehele sample. De respondenten die hierdoor uitvallen, lijken niet door demografische of politieke variabelen (stemgedrag) beïnvloed te zijn.

## Regressieanalyse

### Lasso-procedure

Stel dat de regressievergelijking  $y = X\beta + \varepsilon$  wordt geschat, waarbij vector  $y \in \mathbb{R}^n$  en matrix  $X \in \mathbb{R}^{n \times p}$ .

Doorgaans worden de coëfficiënten in  $\beta$  zodanig geschat, dat de (gekwadrateerde) verschillen tussen de schattingen  $\hat{y}_i$  en de bijbehorende waarnemingen  $y_i$  zo klein mogelijk zijn. Iedere geschatte coëfficiënt  $\beta_j$  heeft een bepaald betrouwbaarheidsinterval. Soms is het wenselijk om variabelen met een kleine coëfficiënt weg te laten. Het model wordt dan overzichtelijker en ook verkleint dit het versturende effect van verbanden tussen regressoren (multicollineariteit) op de schatting van de betrouwbaarheidsintervallen.

De lasso-regressie maakt de coëfficiëntschatting tegelijkertijd met de hiervoor genoemde selectie van coëfficiënten. Bovenop het gekwadrateerde verschil wordt een strafterm geteld die groter is naarmate de coëfficiënt verder afwijkt van nul. Bij een strafcoëfficiënt die groot genoeg is, zijn coëfficiënten automatisch nul, zodat de strafterm geen effect heeft op de doelfunctie. In formulevorm wordt het volgende optimaliseringsprobleem opgelost:

$$\min_{\beta} \sum_{i=1}^n (y - X\beta)_i^2 + \lambda \sum_{j=1}^p |v_j \beta_j|,$$

met  $\lambda > 0$  de zogenoemde algemene strafcoëfficiënt en  $v_j \geq 0$  de strafcoëfficiënt van  $\beta_j$ . De standaardwaarde voor  $\lambda$  volgt uit:

$$\lambda = \lambda_0 \sigma^2 = 2c\sqrt{n}\Phi^{-1}\left(1 - \frac{\gamma}{\log(n_{\text{cluster}})} \frac{1}{2p}\right) \sigma^2 \quad c = 1,1 \quad \gamma = 0,1$$

waarbij  $n_{\text{cluster}}$  het aantal clusters is en  $\sigma^2$  een schatting van de variantie van het residu  $e = y - X\beta$ . Een coëfficiënt  $v_j = 0$  betekent dat de coëfficiënt  $\beta_j$  met OLS wordt bepaald. Deze 'ongestrafte coëfficiënten'

<sup>10</sup> We hebben de stabiliteit van dit model ook getest, gebruikmakend van een zogenoemde ESEM (Exploratory Structural Equation Model (Asparouhov en Muthén 2009; Marsh et al. 2010; Marsh, Morin, Parker en Kaur 2014; Marsh, Muthén, Asparouhov, Lüdtke en Robitzsch 2009; Morin, Marsh en Nagengast 2013)). De resultaten zijn vergelijkbaar en kunnen zodoende als een robuustheidstest van ons confirmatieve model gezien worden.



worden dan verder buiten de lasso-regressie gelaten ('partialing out'). Een grote coëfficiënt  $v_j$  zorgt ervoor dat de bijbehorende coëfficiënt  $\beta_j$  richting nul wordt gedrukt.

De lasso-regressie is uitgevoerd door in Stata het `lasso`-commando te gebruiken, zodat mogelijk zijn:

1. Weging van waarnemingen: De waarnemingen zijn gewogen met dezelfde weegfactor als in de enquête. De weegfactor is gebaseerd op geslacht, opleidingsniveau en leeftijdsgroep.
2. Clustering van residuen: De residuen zijn geclusterd voor groepen met zowel hetzelfde geslacht, hetzelfde opleidingsniveau als dezelfde leeftijdsgroep.
3. Robuuste schattingen voor heteroscedasticiteit: Robuuste schattingen worden verkregen met de strafcoëfficiënten

$$v_j = \sqrt{\frac{\mathbb{E}_i[\tilde{x}_j^2 e_i^2]}{\mathbb{E}_i[e_i^2]}}$$

waarbij  $\tilde{x}_j$  een 'demeaned' regressor is (dus  $\tilde{x}_j = x_j - \mathbb{E}[x_j]$ , zodat  $\mathbb{E}[\tilde{x}_j] = 0$ ) en  $e_i$  het residu van waarneming  $i$ . Om de schattingen cluster-robust te maken

$$v_j = \sqrt{\frac{\mathbb{E}[\tilde{x}_{ci}^2 e_{ci}^2]}{\mathbb{E}[e_{ci}^2]}}$$

waarbij  $\tilde{x}_{ci}$  de 'demeaned' regressor  $i$  van cluster  $c$  is (dus  $\tilde{x}_{ci} = x_{ci} - \mathbb{E}[x_{ci}]$  zodat  $\mathbb{E}[\tilde{x}_{ci}] = 0$ ).

#### Bepaling extreme groepen

De verschillen in gemiddelde scores per dimensie zijn tussen de groepen op basis van één kenmerk vrij beperkt. Dit zijn univariate analyses, want groepen met meerdere kenmerken zijn daarbij niet meegenomen. Het kan echter zijn dat een combinatie van kenmerken andere waarden geeft voor de dimensies. Van de vijf kenmerken die opgenomen zijn in de lasso-regressies – geslacht, leeftijdscategorie, opleidingsniveau, mate van stedelijkheid en herkomst – is daarom bekeken of er een gezamenlijk effect is op één van de dimensies. De kenmerken kunnen als volgt worden omschreven:

Geslacht (2): Man (49,6%), Vrouw (50,4%)

Leeftijdscategorie (4): 15-29 jaar (22,6%), 30-44 jaar (22,2%), 45-64 jaar (33,4%), 65 jaar en ouder (21,8%)

Opleidingsniveau (3): Laag (31,7%), Middel (39,2%), Hoog (29,1%)

Verstedelijking (3): Niet/weinig stedelijk (34,7%), Matig stedelijk (22,4%), (Zeer) sterk stedelijk (42,9%)

Migrantenstatus (2): Nederlandse achtergrond (82%), Migratieachtergrond (18%)

Combinaties van de vijf kenmerken zijn geanalyseerd waarbij aan de volgende twee criteria is voldaan:

1. maximaal drie verschillende kenmerken tegelijk, zodat de groep eenvoudig is te omschrijven;
2. minimaal honderd waarnemingen die aan de kenmerken voldoen, zodat de groep enigszins betrouwbaar is geschat.

Deze procedure is toegepast voor alle veertien dimensies voor zowel typerend als verbondenheid. In totaal zijn er dus 28 plusgroepen die het hoogst scoren op een dimensie, 28 mingroepen die het laagst scoren op een dimensie en 28 verschillen tussen de plusgroep en de bijbehorende mingroep. De verschillen in de tabellen A.1 en A.2 in de bijlage van hoofdstuk 7 tussen de extreme groepen zijn aanmerkelijk groter dan de verschillen op basis van één kenmerk.

#### Mahalanobis-afstand

De kansen die opgenomen zijn in tabel B.1 van de bijlage bij hoofdstuk 7 zijn gebaseerd op een logistische regressieanalyse. De afhankelijke variabele is gelijk aan 1 wanneer iemand tot de groep behoort die zich 'dichtbij' de doorsneesrespondent bevindt. Deze bivariate variabele is gelijk aan 0 wanneer dit niet het geval is. We voegen verschillende sociodemografische variabelen toe als onafhankelijke variabelen. Tabel 4 geeft het resultaat van deze regressieanalyse.

Tabel 4

**Logistische regressie op verbondenheid i.v.m. Mahalanobis-procedure**

Number of strata = 1	Nummer of obs = 3.596		
Number of PSUs = 3.596	DF = 3.595		
N. of poststrata = 36	F (26, 3570) = 2,09		
	Prob > F = 0		
	Mediaan als ijkpunt		
	Odds Ratio	Std. Err.	P> t
man	1,00	(basisgroep)	
vrouw	1,11	0,10	0,23
laag opgeleid	1,00	(basisgroep)	
middelbaar opgeleid	1,12	0,12	0,29
hoog opgeleid	1,20	0,14	0,11
15-29 jaar	1,00	(basisgroep)	
30-44 jaar	0,99	0,18	0,96
45-64 jaar	1,24	0,21	0,21
≥ 65 jaar	1,26	0,26	0,25
t/m 1000 euro	1,00	(basisgroep)	
1001 t/m 2000 euro	1,00	0,11	0,99
meer dan 2000 euro	1,15	0,16	0,30
Nederlandse achtergrond	1,00	(basisgroep)	
migratieachtergrond	0,66	0,08	0,00
niet/weinig stedelijk	1,00	(basisgroep)	
matig stedelijk	0,88	0,09	0,22
(zeer) sterk stedelijk	0,88	0,09	0,22
vrijwilligerswerk	1,00	(basisgroep)	
student	1,68	0,55	0,11
huishouding	1,90	0,54	0,02
pensioen/uitkering	1,90	0,48	0,01
werkend	1,76	0,46	0,03

Gelderland	1,39	0,24	0,06
Utrecht	1,10	0,22	0,64
Noord-Holland	1,30	0,23	0,14
Zuid-Holland	1,20	0,20	0,28
Zeeland	1,78	0,55	0,06
Noord-Brabant	1,17	0,20	0,36
constante	0,34	0,13	0,01

Bron: SCP, DAN'18

Noot: de resultaten zijn op een logistische regressieanalyse gebaseerd. Regressie gebruikt de mediaan als ijkpunt en geeft respondenten die onder de mediaan vallen een 1.

## Hoofdstuk 8

### 2.1 Archetypenanalyse-methode

Evenals factoranalyse is archetypenanalyse een techniek die veelal gebruikt wordt om de dimensionaliteit van een dataset te onderzoeken en te reduceren (Cutler en Breiman 1994). Ook hier kunnen echter de latente factoren – *archetypen* genoemd – interessante theoretische inzichten opleveren. In het SCR wordt archetypenanalyse in combinatie met factoranalyse gebruikt, omdat beide technieken vanuit verschillende aannames en veronderstellingen vertrekken. Zodoende kunnen de technieken samen uiteenlopende informatie uit een dataset naar voren laten komen. Archetypenanalyse is een methode die de extreme posities weergeeft. Hieronder volgt een korte uiteenzetting van de techniek.

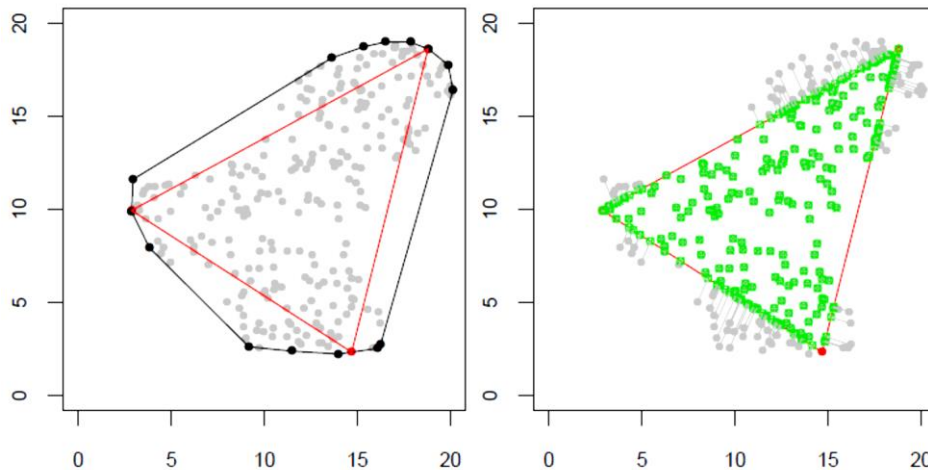
Archetypenanalyse (AA) vertrekt vanuit het idee dat we patronen in een dataset kunnen onderscheiden en definiëren. Deze zogenoemde archetypen zijn primitieve of typerende voorbeelden van objecten (Midgley en Venaik 2012). Ze worden ook wel gemeenplaatsen genoemd. We kunnen denken aan bijvoorbeeld de Held of de Moeder. Doordat de archetypen gedefinieerd worden aan de hand van typen die we vinden aan de uiteinden van de distributie (de extremen, of weberiaanse ideaalbeelden) blijven ze niet mythisch, maar zijn ze reëel. Dit in tegenstelling tot factoranalyse, die latente factoren extraheert aan de hand van gemiddelde typen.

De extreme typen worden vastgesteld via het convexe omhulsel van de verzameling datapunten. Stel de data voor als een puntenwolk in een multidimensionale ruimte. Deze wolk kan omhuld worden door een hyperoppervlak dat alle uiteinden van de multidimensionale wolk omvat. AA maakt gebruik van dit hyperoppervlak om een  $x$ -aantal extreme types vast te stellen (Midgley en Venaik 2012). We bespreken een hypothetisch voorbeeld om de procedure verder inzichtelijk te maken (zie figuur 6). De figuur aan de linkerkant laat een puntenwolk zien. Het hyperoppervlak is aangeduid met een zwarte lijn, de archetypische oplossing met een rode lijn. De drie punten duiden het optimale aantal archetypen aan. Dit doet het algoritme door de afstand van elke datapunt tot de drie archetypen – de zogenoemde *residual sum of squares* (RSS) – te minimaliseren. Elke datapunt krijgt dus een score toegekend op elk archetype en kan zodoende als een composiet gezien worden.<sup>11</sup> De figuur aan de rechterkant illustreert dit door de grijze lijnen die de afstand weergeven tussen de punten die buiten de oplossing vallen en de archetypische oplossing. Het is deze afstand die de optimale oplossing minimaliseert.

<sup>11</sup> De scores zelf liggen altijd tussen de 0 en 1 en tellen gezamenlijk op tot 1. Voor een drie-archetypenmodel betekent dit dus dat:  
 $\text{archetype 1} + \text{archetype 2} + \text{archetype 3} = 1$ .

Figuur 6

Visualisatie van een model met drie archetypen



Figuur afkomstig uit Eugster en Leisch (2009).

De optimale oplossing wordt vastgesteld door middel van een grafische analyse van de RSS ten opzichte van het aantal archetypen. Evenals bij een exploratieve factoranalyse wordt het aantal archetypen gekozen op basis van de knik in de grafiek (*scree plot*). Dit is echter niet alleen van belang. Van even grote waarde is de interpretatie van de archetypen op basis van de parameters. Zodra het aantal archetypen is vastgesteld, kunnen archetypische scores aan elk individu in de dataset toegekend worden.

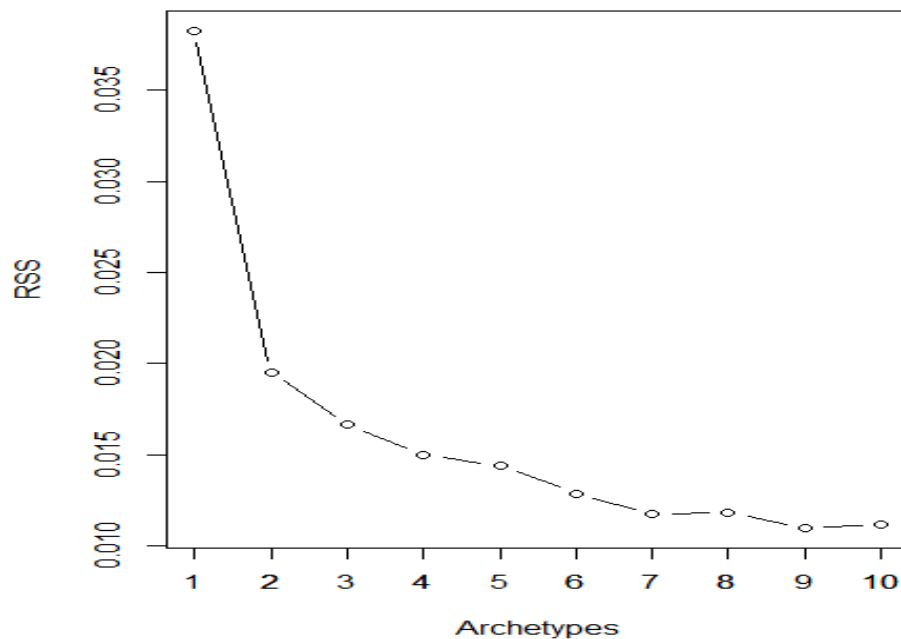
Binnen het SCR wordt AA gebruikt na de exploratieve factoranalyses. We gebruiken het *archetypes package* in R om de analyses uit te voeren (Eugster en Leisch 2009). Als input voor deze analyse gebruiken we het factormodel dat de Nederlandse verbondenheid weergeeft (tabel 7.5 in hoofdstuk 7). Omdat we geïnteresseerd zijn in de vraag op welke manier men zich verbonden voelt met Nederland, gebruiken we de gemiddelde dimensiescores als input voor onze analyse. Deze resultaten zijn gepresenteerd in hoofdstuk 8.

## 2.2 Een typologie van Nederlandse identificatie

We gebruiken het *archetypes package* in R met als input de veertien geschatte dimensies op *verbondenheid*. We gebruiken de gemiddelde score per dimensie. 4019 respondenten hebben voor alle dimensies een score gekregen en kunnen daarom meegenomen worden in de archetypenanalyse (niet-getoonde analyses laten zien dat de relatie tussen sociodemografische kenmerken als leeftijd, stedelijkheid, migrantenstatus en bijvoorbeeld inkomen niet of slechts zeer beperkt samenhangt met de kans om tot de sample van 4019 te behoren). We schatten de optimale oplossingen voor modellen met één tot tien archetypen. Voor elk model laten we het algoritme drie keer draaien om lokale optima te voorkomen. Om een model te kiezen gebruiken we weer Cattell's scree test. De *scree plot* is weergegeven in figuur 7.

Figuur 7

Scree plot archetypenanalyse



We zien dat de RSS het meest afneemt na toevoeging van het tweede archetype. Hierna neemt de RSS gestaag af, totdat deze na zeven factoren bijna constant blijft. Gebaseerd op deze *scree plot* kiezen we ervoor om een archetypenmodel te schatten met twee, drie én vier archetypen. Figuur 8 geeft de oplossingen van de drie modellen grafisch weer. Per dimensie wordt een parameter geschat die laat zien hoe belangrijk die dimensie is voor het respectievelijke archetype. Dit wordt gedaan op de vijfpuntsschaal. Een score van vijf betekent dat de dimensie erg belangrijk is voor het archetype, een score van 1 betekent dat de dimensie niet belangrijk is. We vergelijken de drie modellen op inhoud om vervolgens een keuze te kunnen maken voor ons definitieve model.

Het model met twee archetypen laat zien dat er twee extreme types zijn, waarvan er één lage ladingen op elke dimensie heeft en de ander een relatief hoge lading op alle dimensies behalve de dimensies *ondeugden*, *deugden*, *joods-christelijke traditie* en *islamitische traditie*. Archetype II scoort hoog op de zogenoemde etnische dimensies, zoals *symbolen* en *tradities*, *kunst* en *beroemdheden*. 39% van de respondenten identificeert zich met Archetype I, 61% met Archetype II. De standaarddeviatie is voor beide archetypen gelijk aan 18 procentpunt.

Het drie-archetypenmodel voegt een derde type toe aan het twee-archetypenmodel. Archetypen 1 en 2 komen overeen met Archetypen I en II van het model met twee archetypen. Archetype 3 onderscheidt zich van de twee andere profielen door een voorkeur uit te spreken voor de dimensies die burgerlijke vrijheden onderschrijven. De dimensies *vrijheden en democratie*, *verzorgingsstaat*, en de religieuze dimensies zoals *islamitische traditie* en de *joods-christelijke traditie* worden door dit profiel hoog gewaardeerd. Dit model werken we verder uit in hoofdstuk 8 en vormt de basis voor de analyses die in dat hoofdstuk uitgevoerd worden. Dit omdat dit volgens ons een simpel en intuïtief model is dat de verschillende posities ten opzichte van identificatie met Nederland goed weergeeft. Profiel 1 scoort gemiddeld 34% (met een standaardafwijking van 19 procentpunt). Profiel 2 krijgt een score van 46% (met een standaardafwijking van 21 procentpunt). De gemiddelde score op het derde archetype is het laagst met 20% (en een standaardafwijking van 15 procentpunt).

Het model met vier archetypen splitst het symbolen-en-traditiesprofiel uit in twee profielen, archetypen C en D. Profiel D komt overeen met Archetype II en Archetype 2 uit de voorgaande modellen. Het beschrijft de extreme positie waarin alle dimensies relatief hoog scoren, behalve de dimensies *ondeugden*, *deugden*, *joods-christelijke traditie* en *islamitische traditie*. Profiel D scoort ook laag op deze dimensies, maar

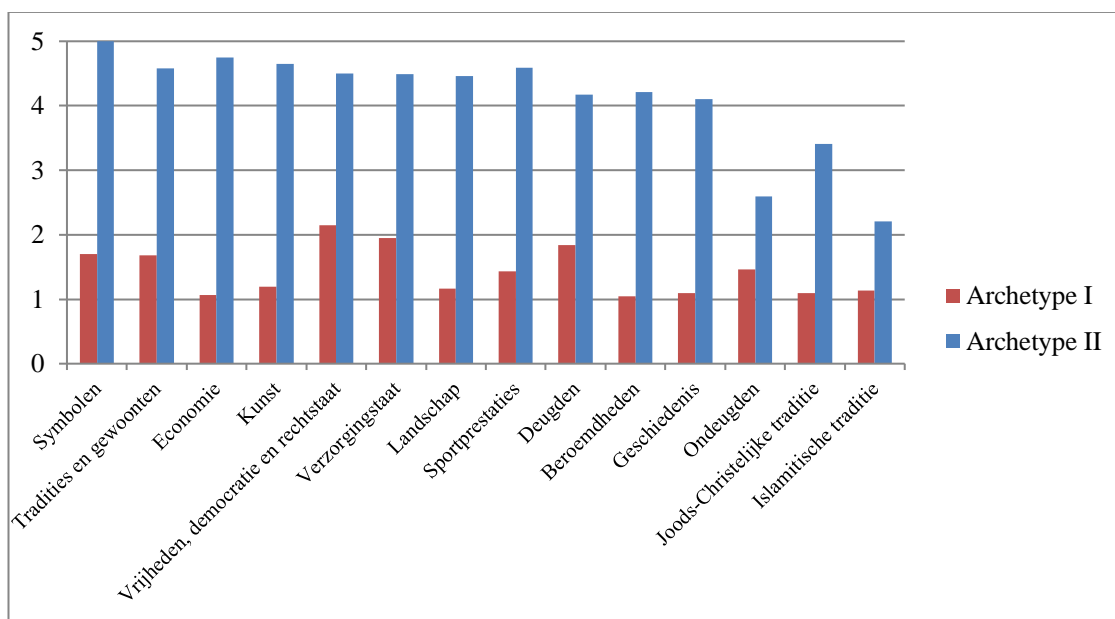
daarnaast relatief gezien hoog op de etnische dimensies, zoals *symbolen, traditie en gewoonten, landschap, economie, beroemdheden* en *geschiedenis*. Profiel A correspondeert met 29% van de respondenten (standaarddeviatie van 20 procentpunt). 14% van de respondenten lijkt op een Profiel B (met een standaarddeviatie van 15 procentpunt). Profiel C blijkt de preferenties van 21% te vatten (standaarddeviatie van 18 procentpunt). Profiel D omschrijft dan de rest van de respondenten, wat neerkomt op 35% van de respondenten (standaarddeviatie van 21 procentpunt).

### 2.3 Robuustheid model

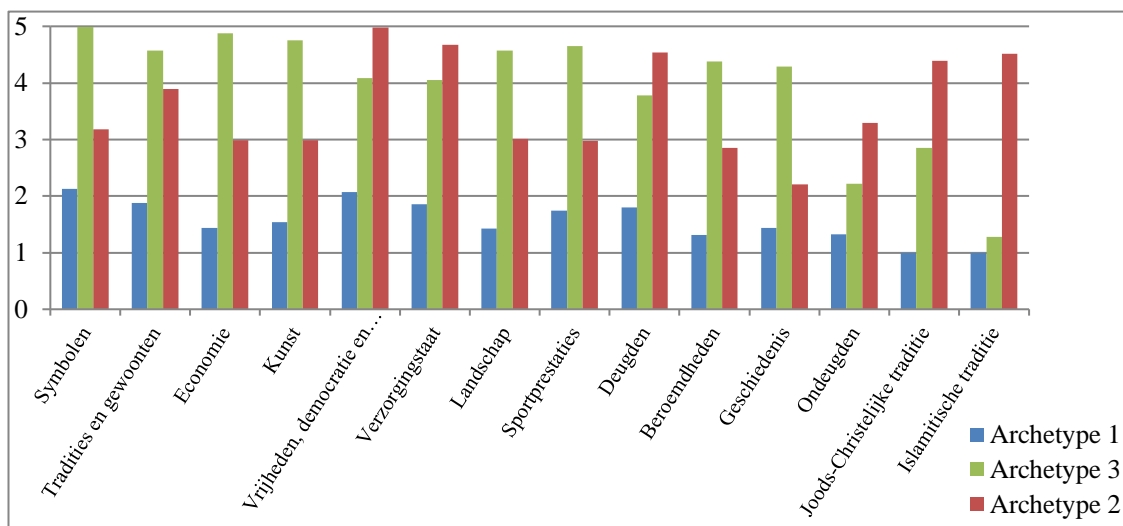
Om zeker te zijn dat we de juiste patronen hebben opgepikt, hebben we ook de twee hoogst ladende items per dimensie geselecteerd en deze als input voor de archytypenanalyse gebruikt. Gebruikmakend van dit criterium selecteren we de items die in tabel A3 (opgenomen in de appendix aan het einde van deze bijlage) weergegeven zijn. De analyses die gebaseerd zijn op deze input geven dezelfde patronen als de resultaten die in hoofdstuk 8 gepresenteerd zijn. Overigens geeft een analyse waarin niet enkel de items met de twee hoogste ladingen worden meegenomen, maar alle items, ook hetzelfde resultaat. Omdat het om zoveel items gaat, is dit echter visueel niet aantrekkelijk en lastiger te interpreteren.

Figuur 8

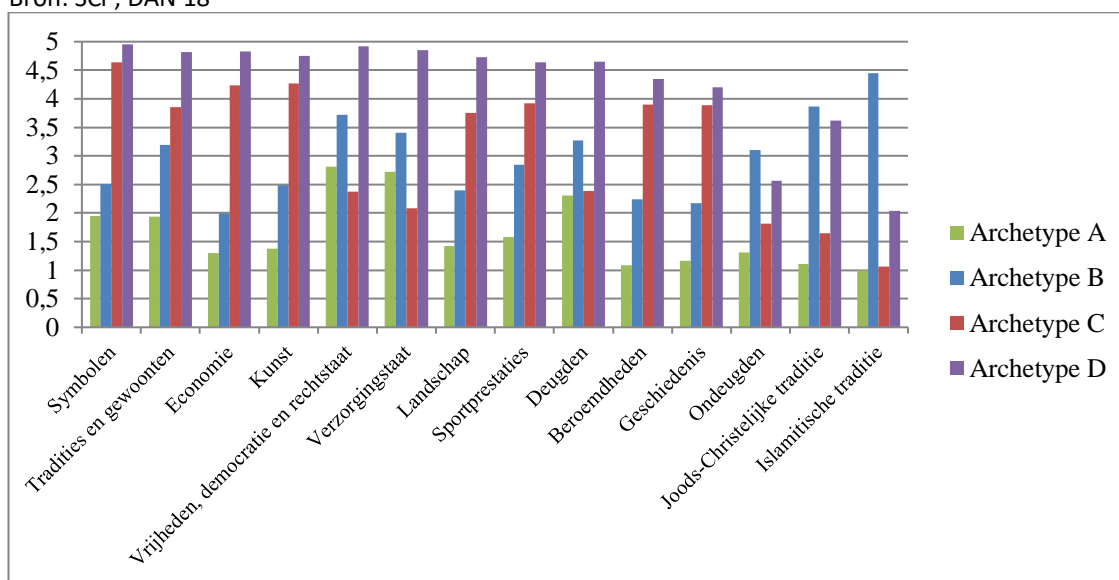
Drie archytypenmodellen



Bron: SCP, DAN'18



Bron: SCP, DAN'18



Bron: SCP, DAN'18

## Conclusie

Deze bijlage heeft een overzicht gegeven van de verschillende methoden die in hoofdstuk 7 en 8 van het SCR 2019 gebruikt zijn. We hebben de achtergrond van exploratieve en confirmatieve factoranalyse gegeven, evenals de onderliggende methode van de archytypenanalyse. Ook de data en de resultaten zijn uitgebreid beschreven om een zo volledig mogelijk beeld te geven van alle stappen die doorlopen zijn. Samen met de beschrijving van de onderzoekopzet in hoofdstuk 18 geeft deze bijlage inzicht in de aard van de analyses.

De opzet van het onderzoek heeft in grote mate de methodologie bepaald. Zo zijn de exploratieve analyses niet alleen op het LISS-panel uitgevoerd, maar ook als een robuustheidsanalyse op het I&O-panel. Alhoewel de twee enquêtes verschillen, blijken de uitkomsten van de exploratieve analyses in grote mate te overlappen. Dit suggereert dat de metingen die in het SCR zijn opgenomen van de cognitieve en affectieve aspecten van Nederlandse identiteit stabiel en valide zijn. Ook de tests van het confirmatieve model middels een ESEM onderschrijven dit.

De uiteenlopende aannames die aan de factoranalyses en archytypenanalyses ten grondslag liggen, resulteren in verschillende inzichten die elkaar aanvullen. Ze beantwoorden ook andere vragen. Factoranalyse onderzoekt de verschillende dimensies van identificatie met Nederland, welke deze zijn en hoe deze te

interpreteren zijn. Archetypenanalyse gebruikt deze dimensies om uit te zoeken hoe de Nederlander zich met Nederland verbindt. Hieraan blijken drie profielen ten grondslag te liggen, die als configuratie de veertien dimensies van de verbondenheid met Nederland samenvatten.

## Referenties

- Asparouhov, T. en B. Muthén (2009). Exploratory structural equation modeling. In: *Structural Equation Modeling*, jg. 16, nr. 3, p. 397-438.
- Asparouhov, T. en B. Muthén (2018). *SRMR in Mplus*. Technical report 2 May 2018.
- Cutler, A. en L. Breiman (1994). Archetypal analysis. In: *Technometrics*, jg. 36, nr. 4, p. 338-347.
- Enders, C.K. (2001). A primer on maximum likelihood algorithms available for use with missing data. In: *Structural Equation Modeling*, jg. 8, nr. 1, p. 128-141.
- Eugster, M.J.A. en F. Leisch (2009). From Spider-Man to Hero – Archetypal Analysis in R. In: *Journal of Statistical Software*, jg. 30, nr. 8, p. 1-23.
- Hu, L. en P.M. Bentler (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. In: *Structural Equation Modeling*, jg. 6, nr. 1, p. 1-55.
- Kenny, D.A. (2014). Measuring Model Fit. <http://davidakenny.net/cm/fit.htm>.
- Marsh, H.W., O. Lüdtke, B. Muthén, R. Asparouhov, A.J.S. Morin en U. Trautwein (2010). A new look at the Big Five factor structure through exploratory structural equation modeling. In: *Psychol. Assess.*, jg. 22, nr. 3, p. 471-491.
- Marsh, H.W., A.J.S. Morin, P.D. Parker en G. Kaur (2014). Exploratory Structural Equation Modeling: An Integration of the Best Features of Exploratory and Confirmatory Factor Analysis. In: *Annu. Rev. Clin. Psychol.*, jg. 10, nr. 1, p. 85-110.
- Marsh, H.W., B. Muthén, T. Asparouhov, O. Lüdtke, A. Robitzsch, A.J.S. Morin en U. Trautwein (2009). Exploratory structural equation modeling, integrating CFA and EFA: application to students' evaluations of university teaching. In: *Structural Equation Modeling*, jg. 16, nr. 3, p. 439-476.
- Midgley, D. en S. Venaik (2012). Marketing management in MNC subsidiaries: An archetypal analysis of heterogeneity in strategy and organization. *INSEAD faculty & research working paper 72*.
- Morin, A.J.S., H.W. Marsh en B. Nagengast (2013). Exploratory structural equation modeling: an introduction. In: G.R. Hancock en R.O. Mueller (red.), *Structural Equation Modeling: A Second Course* (p. 395-436). Greenwich, CT: IAP.
- Wansbeek, T. en E. Meijer (2000). *Measurement Error and Latent Variables in Econometrics*. Amsterdam: North Holland.



## Appendix

Tabel A1

Output EFA en CFA op typerend

	EFA Split sample	SEM	Cronbach's alfa <sup>†</sup>
<b>vrijheden, democratie en rechtstaat</b>			0,91
vrijheid van meningsuiting	0,699	0,740	
vrijheid van geloof	0,599	0,694	
vrijheid van demonstratie	0,673	0,694	
vrijheid van pers	0,730	0,743	
vrijheid van onderwijs	0,606	0,693	
gelijkheid van man en vrouw (voor de wet)	0,646	0,705	
gelijkheid van homo en hetero (voor de wet)	0,517	0,592	
de democratie	0,732	0,732	
de Grondwet	0,621	0,674	
het algemeen kiesrecht	0,752	0,735	
scheiding van kerk en staat	0,502	0,556	
abortuswetgeving	0,437	0,522	
<b>ondeugden</b>			0,64
corruptie	0,491		
achterkamertjespolitiek	0,480		
gedoogbeleid			
bothed	0,490	0,639	
betweterig zijn	0,496	0,664	
individualisme			
klagen	0,411	0,463	
<b>deugden</b>			0,81
internationale oriëntatie	0,392		
tolerantie	0,449	0,483	
ruimdenkendheid	0,476	0,512	
bescheidenheid	0,448	0,443	
een nuchtere, praktische houding	0,406	0,453	
'gewoon' doen; niet naast je schoenen lopen	0,386	0,437	
gezelligheid	0,453	0,569	
netheid	0,460	0,515	
gastvrijheid <sup>1</sup>	0,533	0,543	
vreedzaamheid	0,481	0,617	
zorgzaamheid <sup>1</sup>	0,513	0,618	
elegantie	0,413	0,409	
zelfbeheersing	0,509	0,518	
	EFA Split sample	SEM	Cronbach's alfa <sup>†</sup>

<b>tradities en gewoonten</b>			0,89
dodenherdenking (4 mei)	0,497	0,624	
Bevrijdingsdag (5 mei)	0,483	0,606	
Koningsdag	0,508	0,657	
fietsen	0,394	0,586	
Prinsjesdag	0,537	0,594	
pakjesavond en sinterklaas	0,542	0,669	
beschuit met muisjes bij geboorte kind	0,581	0,672	
bij elkaar op de koffie gaan (met een koekje) <sup>1</sup>	0,472	0,530	
1 aprilgrap			
Elfstedentocht	0,505	0,598	
boekentas met vlag uit het raam bij diploma middelbare school	0,541	0,632	
haring happen	0,540	0,593	
schaatsen en ijspret	0,493	0,602	
oliebollen en appelflappen tijdens oud en nieuw	0,632	0,671	
stampot	0,501	0,611	
eten prakken	0,407		
<b>directheid/mengelmoes</b>			
directheid in communicatie (bijv. tutoyeren)	0,396		
coffeeshops	0,402		
botheid	0,342		
gedoogbeleid	0,328		
niet godsdienstig zijn	0,320		
<b>Islamitische traditie</b>			0,84
Ramadan	0,911	0,884	
Suikerfeest	0,884	0,872	
islam	0,591	0,614	
hoofddoek	0,553	0,594	
<b>sportprestaties</b>			0,89
EK voetbal 1988	0,770	0,771	
WK voetbal 1974	0,785	0,768	
prestaties Nederlandse sporters	0,669	0,737	
oranjegekte <sup>1</sup>	0,383	0,522	
oranjegevoel	0,396	0,544	
Olympische medailles Ireen Wüst	0,710	0,768	
winst Grand Prix Spanje Max Verstappen	0,663	0,684	
Olympische medailles Fanny Blankers-Koen	0,664	0,701	
Giro-winst Tom Dumoulin	0,714	0,720	

	EFA Split sample	SEM	Cronbach's alfa <sup>†</sup>
<b>verzorgingsstaat</b>			0,78
de verzorgingsstaat	0,569	0,642	
het pensioenstelsel	0,617	0,651	
de gezondheidszorg	0,594	0,698	
het onderwijsstelsel	0,556	0,675	
de woningcorporaties	0,506	0,527	
<b>Joods-christelijke traditie</b>			0,78
rooms-katholicisme	0,724	0,764	
protestantisme <sup>1</sup>	0,766	0,762	
boeddhisme			
jodendom	0,496	0,565	
humanisme	0,382	0,448	
kerken	0,670	0,703	
<b>geschiedenis</b>			0,78
de Tweede Wereldoorlog	0,504	0,571	
de Tachtigjarige Oorlog	0,572	0,577	
de Watersnoodramp van 1953	0,469	0,608	
de Gouden Eeuw	0,450	0,546	
de Koude Oorlog	0,419	0,415	
het slavernijverleden van Nederland	0,457	0,515	
de moord op Pim Fortuyn	0,388	0,477	
de VOC	0,444	0,524	
de NSB	0,444	0,489	
<b>symbolen</b>			0,85
de Nederlandse taal	0,441	0,595	
het Koninklijk Huis	0,381	0,586	
de Nederlandse vlag	0,498	0,666	
het Nederlandse volkslied – het Wilhelmus	0,486	0,610	
molens	0,605	0,705	
tulpen	0,532	0,666	
klompen	0,548	0,585	
kaas	0,519	0,642	
kleur oranje	0,491	0,640	
<b>beroemdheden</b>			0,88
André Kuipers	0,551	0,693	
Johan Cruijff	0,421	0,638	
André Hazes	0,547	0,609	
Armin van Buuren	0,728	0,718	
Ayaan Hirsi Ali	0,401		
Mies Bouwman	0,466	0,618	
Epke Zonderland	0,671	0,740	

	EFA Split sample	SEM	Cronbach's alfa <sup>†</sup>
Anky van Grunsven	0,673	0,721	
Pim Fortuyn	0,542	0,585	
Ahmed Aboutaleb	0,367		
Doutzen Kroes	0,707	0,676	
Nederlandse DJ's	0,506	0,515	
<b>oude beroemdheden</b>			
Willem van Oranje	0,694		
Michiel de Ruyter	0,807		
Piet Hein	0,747		
<b>landschap</b>			
			0,88
polderlandschap	0,476	0,636	
heidelandschap	0,590	0,658	
dijken	0,412	0,599	
rivieren	0,705	0,650	
weilanden	0,695	0,712	
dichtbevolkt <sup>1</sup>	0,477	0,477	
duinen, stranden en zee	0,622	0,693	
bollenvelden			
de Veluwe	0,457	0,620	
Waddeneilanden en Waddenzee	0,438	0,639	
wolkenluchten	0,692	0,645	
de vier seizoenen en het weer	0,649	0,633	
regen	0,637	0,574	
<b>waterbeheer</b>			
			0,73
de Deltawerken	0,522	0,731	
grote Nederlandse internationale bedrijven			
de haven van Rotterdam			
inpoldering Noordoostpolder en Flevoland	0,541	0,736	
Nederlandse land- en tuinbouw			
Gasbel Groningen			
Nederlandse baggeraars	0,568	0,636	
Keukenhof			
de gulden			
<b>kunst</b>			
			0,87
de Nachtwacht van Rembrandt	0,496	0,690	
de Zonnebloemen van Vincent van Gogh	0,507	0,665	
de Efteling	0,411	0,569	
het Rijksmuseum <sup>1</sup>	0,482	0,681	
Soldaat van Oranje – theatershow	0,394	0,545	
Jip en Janneke van Annie M.G. Schmidt	0,661	0,703	
Nijntje van Dick Bruna	0,641	0,682	

	EFA Split sample	SEM	Cronbach's alfa <sup>†</sup>
boeken van Mulisch	0,509	0,559	
het dagboek van Anne Frank	0,436	0,648	
Max Havelaar van Multatuli	0,471	0,564	
RMSEA	0,026	0,035	
Kans dat RMSEA < 0,05	1	1	
CFI	0,887	0,761	
TLI	0,863	0,755	
SRMR	0,020	0,074	

†Alfa's zijn berekend voor de items die in de CFA zijn opgenomen.

Bron: SCP, DAN'18

Tabel A2

Output EFA en CFA op verbondenheid

	exploratief		confirmatief	Cronbach's alfa <sup>†</sup>
	15 factoren	14 factoren	14 factoren	
<b>vrijheden, democratie en rechtstaat</b>				0,91
vrijheid van meningsuiting	0,671	0,659	0,723	
vrijheid van geloof	0,522	0,461	0,615	
vrijheid van demonstratie	0,600	0,602	0,672	
vrijheid van pers	0,658	0,640	0,727	
vrijheid van onderwijs	0,567	0,475	0,653	
gelijkheid van man en vrouw (voor de wet)	0,620	0,625	0,719	
gelijkheid van homo en hetero (voor de wet)	0,493	0,625	0,619	
de democratie	0,701	0,626	0,730	
de Grondwet	0,647	0,564	0,704	
het algemeen kiesrecht	0,664	0,592	0,737	
scheiding van kerk en staat	0,531	0,532	0,584	
abortuswetgeving	0,549	0,644	0,577	
euthanasiewetgeving	0,486	0,589	0,542	
de verzorgingsstaat	0,391			
<b>ondeugden</b>				0,79
corruptie		0,412	0,435	
discriminatie		0,414	0,431	
achterkamertjespolitiek	0,406	0,474	0,560	
botheid	0,596	0,509	0,634	
betweterig zijn	0,572	0,481	0,675	
individualisme	0,457	0,385	0,514	
klagen	0,509	0,441	0,582	
zuinigheid	0,431	0,332	0,484	
luiheid	0,398	0,441	0,460	
de bureaucratie		0,453	0,484	
de NSB		0,400		
<b>deugden</b>				0,87
tolerantie	0,408		0,548	
ruimdenkendheid	0,420		0,594	
bescheidenheid	0,523		0,555	
een nuchtere, praktische houding	0,398		0,533	
'gewoon' doen; niet naast je schoenen lopen	0,390		0,507	
gezelligheid	0,519	0,403	0,610	
netheid	0,571	0,535	0,596	
gastvrijheid	0,621	0,577	0,608	
vreedzaamheid	0,529	0,511	0,656	
zorgzaamheid	0,655	0,632	0,689	
diversiteit	0,386		0,499	

	exploratief 15 factoren	confirmatief 14 factoren	Cronbach's alfa <sup>†</sup> 14 factoren	
de verzorgingsstaat		0,430		
het pensioenstelsel		0,391		
de gezondheidszorg	0,387	0,496		
het onderwijsstelsel		0,437		
<b>verzorgingsstaat</b>				<b>0,81</b>
de verzorgingsstaat			0,754	
het pensioenstelsel			0,680	
de gezondheidszorg			0,777	
het onderwijsstelsel			0,678	
<b>factor X</b>				
internationale oriëntatie		0,416		
ruimdenkendheid		0,374		
een nuchtere, praktische houding		0,489		
'gewoon' doen; niet naast je schoenen lopen		0,427		
<b>tradities en gewoonten</b>				<b>0,91</b>
dodenherdenking (4 mei)	0,443	0,428	0,584	
Bevrijdingsdag (5 mei)	0,466	0,456	0,604	
Koningsdag	0,379	0,374	0,618	
fietsen	0,403	0,411	0,547	
Prinsjesdag	0,409	0,418	0,618	
pakjesavond en sinterklaas	0,562	0,555	0,643	
beschuit met muisjes bij geboorte kind	0,635	0,644	0,711	
bij elkaar op de koffie gaan (met een koekje)	0,476	0,480	0,597	
Elfstedentocht	0,477	0,464	0,614	
1 aprilgrap	0,497	0,511	0,518	
Kerstmis	0,379	0,383	0,581	
boekentas met vlag uit het raam bij diploma middelbare school	0,599	0,609	0,690	
haring happen	0,517	0,524	0,605	
schaatsen en ijspret	0,565	0,557	0,672	
oliebollen en appelflappen tijdens oud en nieuw	0,619	0,628	0,688	
stampot	0,504	0,519	0,598	
eten prakken	0,422	0,440	0,451	
<b>sportprestaties</b>				<b>0,93</b>
EK voetbal 1988	0,805	0,804	0,812	
WK voetbal 1974	0,832	0,831	0,809	
prestaties Nederlandse sporters	0,760	0,754	0,813	
oranjegekte	0,651	0,651	0,754	
oranjegevoel	0,630	0,629	0,759	
Olympische medailles Ireen Wüst	0,748	0,743	0,824	

	exploratief 15 factoren	confirmatief 14 factoren	Cronbach's alfa <sup>†</sup> 14 factoren	
winst Grand Prix Spanje Max Verstappen	0,702	0,702	0,752	
Olympische medailles Fanny Blankers-Koen	0,656	0,653	0,744	
Giro-winst Tom Dumoulin	0,772	0,764	0,758	
<b>Islamitische traditie</b>				0,86
Ramadan	0,885	0,890	0,886	
Suikerfeest	0,862	0,865	0,868	
islam	0,645	0,648	0,708	
hoofddoek	0,572	0,575	0,653	
<b>landschap</b>				0,93
polderlandschap	0,617	0,655	0,774	
heidelandschap	0,696	0,713	0,747	
dijken	0,523	0,567	0,750	
rivieren	0,707	0,727	0,749	
weilanden	0,754	0,773	0,782	
duinen, stranden en zee	0,682	0,702	0,764	
bollenvelden	0,396	0,446	0,714	
de Veluwe	0,632	0,646	0,747	
Waddeneilanden en Waddenzee	0,619	0,644	0,763	
wolkenluchten	0,707	0,711	0,686	
de vier seizoenen en het weer	0,709	0,701	0,655	
regen	0,632	0,630	0,584	
<b>factor Y</b>				
gelijkheid van homo en hetero (voor de wet)	0,414			
euthanasiewetgeving	0,381			
Gay Pride Amsterdam	0,466			
coffeeshops	0,398			
<b>Factor Z</b>				
de NSB	0,359			
files	0,350			
de Koude Oorlog	0,311			
<b>Joods-christelijke traditie</b>				0,84
rooms-katholicisme	0,779	0,782	0,822	
protestantisme	0,787	0,797	0,828	
jodendom	0,688	0,682	0,724	
kerken	0,770	0,782	0,818	
humanisme	0,411	0,389	0,417	
boeddhisme	0,441	0,416	0,424	



	exploratief 15 factoren	confirmatief 14 factoren	Cronbach's alfa <sup>†</sup> 14 factoren	
<b>beroemdheden</b>				0,90
André Kuipers	0,569	0,492	0,726	
Johan Cruijff	0,483	0,490	0,715	
André Hazes	0,523	0,589	0,648	
Armin van Buuren	0,724	0,759	0,746	
Mies Bouwman	0,459	0,404	0,675	
Epke Zonderland	0,656	0,565	0,755	
Anky van Grunsven	0,641	0,550	0,742	
Pim Fortuyn	0,449	0,490	0,575	
Ahmed Aboutaleb	0,399	0,355	0,520	
Doutzen Kroes	0,621	0,618	0,663	
Nederlandse DJ's	0,463	0,587	0,592	
<b>geschiedenis</b>				0,83
de Tweede Wereldoorlog	0,556	0,547	0,633	
de Tachtigjarige oorlog	0,477	0,468	0,636	
de Watersnoodramp van 1953	0,435	0,412	0,693	
de Gouden Eeuw	0,434	0,407	0,696	
de Koude Oorlog	0,466	0,479	0,512	
het slavernijverleden van Nederland	0,472	0,471	0,487	
de moord op Pim Fortuyn	0,462	0,458	0,552	
de VOC	0,470	0,437	0,706	
de NSB	0,403	0,414		
<b>symbolen</b>				0,93
de Nederlandse taal	0,386	0,411	0,511	
het Koninklijk Huis	0,431	0,488	0,666	
de Nederlandse vlag	0,522	0,577	0,715	
het Nederlandse volkslied – het Wilhelmus	0,470	0,537	0,672	
molens	0,424	0,387	0,765	
tulpen	0,434	0,391	0,768	
klompen	0,420	0,360	0,701	
kaas	0,385	0,328	0,672	
kleur oranje	0,541	0,541	0,667	
<b>economie</b>				0,91
de Deltawerken	0,556	0,540	0,764	
grote Nederlandse internationale bedrijven	0,420	0,426	0,650	
de haven van Rotterdam	0,567	0,547	0,780	
inpoldering Noordoostpolder en Flevoland	0,565	0,532	0,754	
Nederlandse land- en tuinbouw	0,447	0,429	0,752	
Nederlandse baggeraars	0,647	0,627	0,750	
Keukenhof	0,390	0,337	0,695	

	exploratief 15 factoren	confirmatief 14 factoren	Cronbach's alfa <sup>†</sup> 14 factoren
<b>kunst</b>			0,90
de Nachtwacht van Rembrandt	0,626	0,641	0,793
de Zonnebloemen van Vincent van Gogh	0,620	0,641	0,758
de Efteling	0,390	0,407	0,568
het Rijksmuseum	0,628	0,647	0,775
Jip en Janneke van Annie M.G. Schmidt	0,583	0,590	0,703
Nijntje van Dick Bruna	0,597	0,602	0,703
boeken van Mulisch	0,502	0,512	0,609
het dagboek van Anne Frank	0,493	0,510	0,714
Max Havelaar van Multatuli	0,531	0,532	0,669
RMSEA	0,029	0,028	0,041
Kans dat RMSEA < 0,05	1	1	1
CFI	0,885	0,894	0,738
TLI	0,862	0,872	0,732
SRMR	0,020	0,019	0,100

<sup>†</sup>Alfa's zijn berekend voor de items die in de CFA zijn opgenomen.

Bron: SCP, DAN'18

### Tabel A3

Overzicht van de twee hoogst ladende items per verbondenheid-dimensie

<b>Items</b>
<b><i>vrijheden, democratie en rechtstaat</i></b> het algemeen kiesrecht de democratie
<b><i>ondeugden</i></b> betweterig zijn botheid
<b><i>deugden</i></b> zorgzaamheid vreedzaamheid
<b><i>verzorgingsstaat</i></b> de gezondheidszorg de verzorgingsstaat
<b><i>tradities en gewoonten</i></b> beschuit met muisjes bij geboorte kind boekentas met vlag uit het raam bij diploma middelbare school
<b><i>sportprestaties</i></b> Olympische medailles Ireen Wüst prestaties Nederlandse sporters
<b><i>Islamitische traditie</i></b> Ramadan Suikerfeest
<b><i>landschap</i></b> weilanden polderlandschap
<b><i>Joods-christelijke traditie</i></b> protestantisme katholicisme
<b><i>beroemdheden</i></b> Epke Zonderland Armin van Buuren
<b><i>geschiedenis</i></b> de VOC de Gouden Eeuw
<b><i>symbolen</i></b> tulpen molens

***economie***

de haven van Rotterdam  
de Deltawerken

***Kunst***

de Nachtwacht van Rembrandt  
het Rijksmuseum

---

Bron: SCP, DAN'18