

CROSS-CULTUREEL ONDERZOEK: METHODOLOGISCHE PROBLEMEN BIJ DE TOEPASSING VAN EEN VERKORTE F-SCHAAL IN FRANSTALIG EN NEDERLANDSTALIG BELGIË

Jerry Welkenhuysen-Gybels (1)

1. INLEIDING EN PROBLEEMSTELLING

In de internationale literatuur wordt steeds meer aandacht besteed aan de methodologische aspecten en problemen die verbonden zijn aan cross-cultureel onderzoek. Dit soort studies wordt, naast de problemen waarmee het intracultureel onderzoek reeds dient af te rekenen, tevens geconfronteerd met een aantal specifieke moeilijkheden die inherent zijn aan de aard van dergelijk onderzoek (Welkenhuysen-Gybels, 1997). De scores van respondenten uit de verschillende culturele groepen die met elkaar vergeleken worden, op een bepaald item of een bepaald latent concept zijn immers niet steeds rechtstreeks met elkaar te vergelijken. De vergelijkbaarheid van dergelijke scores wordt bepaald door hun equivalentieniveau.

Voor een definitie van het begrip (in)equivalentie kunnen we terecht bij Berry e.a. (1992 (1993), p. 238). Zij maken een onderscheid tussen de *comparison scale* en de *measurement scale* van een theoretisch concept. De *comparison scale* of de schaal van het concept is een hypothetische schaal die voor een bepaald concept wordt gepostuleerd. Gezien het hypothetisch karakter van deze schaal, dient voor de meting van het concept echter een *measurement scale* gebruikt te worden. Men spreekt van inequivalentie in de scores van twee of meer culturele groepen indien geobserveerde verschillen tussen deze groepen op de *measurement scale* niet samengaan met corresponderende verschillen op de *comparison scale*. In de literatuur vindt men echter

ook nog andere definities terug van dit begrip. Johnson (1998) vond in de literatuur meer dan vijftig verschillende definities terug.

Van de Vijver & Leung (1997, p. 8-9) onderscheiden drie niveaus van equivalentie. Deze drie niveaus verhouden zich op een hiërarchische wijze tot elkaar. Het eerste en laagste niveau van equivalentie noemen ze *construct equivalence*. Dit houdt in dat het gebruikte instrument hetzelfde latente concept meet in elk van de onderzochte groepen. Een volgend (en hiërarchisch hoger) niveau van equivalentie is *measurement unit equivalence*. Dit wordt bereikt indien het meetinstrument in elke culturele groep op eenzelfde intervalschaal kan worden uitgedrukt (de afstand tussen de meeteenheden van de *measurement scale* moet dus hetzelfde zijn in elke culturele groep). Poortinga (1989) onderscheidt tevens de situatie waarin de afstand tussen de meeteenheden van de *measurement scale* in de ene groep een lineaire combinatie is van deze afstand in de andere groepen. Na het doorvoeren van een lineaire transformatie kan in dit geval dus eveneens het niveau van *measurement unit equivalence* bereikt worden. Dit equivalentieniveau impliceert niet dat de meet-schalen in elke groep eenzelfde nulpunt hebben. Deze laatste voorwaarde is wel noodzakelijk om het laatste (en hoogste) niveau van equivalentie te bereiken, het niveau van de *scalar equivalence* of *full score comparability*. Little (1997) verwijst naar dit niveau met de term *measurement equivalence*. Dit niveau vereist dat het meetinstrument in elke culturele groep op dezelfde ratioschaal kan worden uitgedrukt. Enkel indien *scalar equivalence* werd vastgesteld, kan men de scores van verschillende culturele groepen rechtstreeks met elkaar vergelijken, aldus van de Vijver & Leung (1997). Deze stelling vindt men eveneens terug bij Little (1997, p. 56): "Although measurement non-equivalence can be a meaningful analytical outcome, it disallows, when sufficiently strong, quantitative construct comparisons."

Het equivalentieniveau van de scores die in de verschillende culturele groepen worden bekomen, wordt echter bedreigd door eveneens drie vormen van vertekening (van de Vijver & Leung, 1997, p. 10-18). Een eerste vorm van vertekening is *construct bias*. Deze doet zich onder meer voor wanneer het gemeten latente concept niet identiek is voor alle culturele groepen of indien 'brede' concepten aan de hand van een (te) beperkt aantal items worden bevraagd. In dit laatste geval spreekt men van *construct underrepresentation*. *Method bias* vloeit voort uit de specifieke eigenschappen van het instrument of de afname ervan. Verschillen in antwoordstijlen, zoals het geven van sociaal wenselijke antwoorden of het gebruik van de extremen van een schaal, kunnen leiden tot deze vorm van bias. Ook verschillen in de vertrouwdheid met het instrument en de fysieke omstandigheden van de afname van het instrument alsook communicatieproblemen tussen de interviewer en de respondent vormen een bron van *method bias*. *Item bias*, ten slotte, kan veroorzaakt worden door incidentele verschillen in de geschiktheid van de inhoud van een item of door een inadequate formulering of vertaling van een item. Bij *method* en *item bias* wordt nog een bijkomend onderscheid gemaakt tussen een uniforme en een niet-uniforme bias. Een *uniform bias* beïnvloedt de scores op min of meer dezelfde

wijze ongeacht het niveau van de scores. Bij *nonuniform bias* verschilt de mate van vertekening naargelang van het niveau van de scores.

Het optreden van deze vormen van vertekening heeft uiteraard een invloed op het equivalentieniveau en dus ook op de vergelijkbaarheid van de scores in de verschillende culturele groepen (van de Vijver & Leung, 1997, p. 18-20). Algemeen kan men stellen dat de aanwezigheid van één of meerdere van de hierboven beschreven vormen van bias, het equivalentieniveau van de scores doet dalen. In tabel 1 wordt de invloed van de drie vormen van vertekening op het equivalentieniveau van de scores schematisch weergegeven.

Tabel 1. Wordt het equivalentieniveau beïnvloed door vertekening?^a

Soort bias	Equivalentieniveau		
	Construct	Measurement unit	Scalar
<i>Construct bias</i>	Ja	Ja	Ja
<i>Method bias: uniform</i>	Nee	Nee	Ja
<i>niet-uniform</i>	Nee	Ja	Ja
<i>Item bias: uniform</i>	Nee	Nee	Ja
<i>niet-uniform</i>	Nee	Ja	Ja

^aBron: van de Vijver, 1998, p. 49.

Construct bias maakt iedere vorm van vergelijking tussen de scores onmogelijk. *Method* en *item bias* hebben geen invloed op de construct equivalence. Immers indien dit het hoogste equivalentieniveau is, zijn numerieke vergelijkingen tussen cross-culturele scores sowieso niet toegestaan. Een uniforme bias heeft geen invloed op de *measurement unit equivalence* vermits alle scoreniveaus op min of meer dezelfde wijze worden beïnvloed en omdat ook op dit niveau een rechtstreekse vergelijking van scores niet mogelijk is. Bij een uniforme bias treedt echter wel een verschuiving van het nulpunt op. Dit heeft tot gevolg dat men geen aanspraak meer kan maken op *scalar equivalence*. Indien een niet-uniforme bias optreedt, is de afstand tussen de meeteenheden van een schaal niet meer gelijk voor elk van de onderzochte groepen. *Construct equivalence* is dan bijgevolg het hoogst mogelijke equivalentieniveau.

In deze bijdrage zullen we nagaan in welke mate de scores van Franstalige en Nederlandstalige respondenten in België op een verkorte autoritarismeschaal met elkaar vergelijkbaar zijn. De centrale stelling van van de Vijver & Leung (1997) is immers dat de equivalentie van de scores in de onderzochte culturele groepen niet als vanzelfsprekend mag worden aangenomen, maar steeds dient te worden aangetoond. We zullen daarom nagaan of eventuele verschillen die gevonden worden

tussen Franstalige en Nederlandstalige respondenten te wijten zijn aan valide cross-culturele verschillen dan wel of deze verschillen (mede) veroorzaakt worden door *construct* of *item bias*. Bij gebrek aan gegevens kunnen de data niet onderzocht worden op de aanwezigheid van *method bias*. De aanwezigheid van deze vorm van vertekening is echter niet uit te sluiten. Alleen al bijvoorbeeld het feit dat de data voor beide groepen werden verzameld door verschillende onderzoeksbureaus (door ISPO voor de Nederlandstalige en PIOP voor de Franstalige respondenten) zou kunnen zorgen voor een vertekening in de resultaten.

2. MEETINSTRUMENT EN POPULATIE

2.1. De verkorte autoritarismeschaal

Autoritarisme werd door Adorno e.a. (1950 (1982)) als een persoonlijkheidskenmerk beschouwd. De verklaring voor dit kenmerk zochten zij bij de psycho-analyse. Autoritaire persoonlijkheden zouden worden gekenmerkt door een betrekkelijk zwak ontwikkeld *Ich* dat in onvoldoende mate in staat is om de maatschappelijk ontoelaatbare impulsen vanuit het *Es* te verdringen. Ter compensatie voor dit zwakke *Ich* zouden autoritaire personen een beroep doen op een sterk, maar geëxternaliseerd *Über-Ich* ten einde de verdringing van de driften vanuit het *Es* alsnog te realiseren. "Het klassieke concept van Adorno e.a. (1950 (1982)) verwijst naar een cluster van negen zogenaamde subsyndromen waaronder het zich strikt houden aan conventionele waarden en normen (*conventionalisme*), een onkritische onderwerping aan en een onkritische houding tegenover morele autoriteiten (*autoritaire submissie*), en een autoritaire agressie tegenover normovertreders als cruciaal beschouwd mogen worden" (Billiet, 1993, p. 154). Andere auteurs geven echter argumenten voor een sociologische verklaring van het autoritarisme (zie onder meer: Scheepers & Eisinga, 1991).

Voor het onderzoek in deze bijdrage zal een verkorte versie van de oorspronkelijke autoritarisme- of Fascismeschaal (kortweg: F-schaal) worden gebruikt. Deze bestond zowel voor de Nederlandstalige als de Franstalige respondenten uit de steekproef uit 7 items en werd ontleend aan Meloen (1983). Op één na (item 7) zijn deze items een bijna letterlijke vertaling van de oorspronkelijke items die door Adorno e.a. (1950 (1982)) werden opgesteld (zie bijlage). Alle items zijn in de oorspronkelijke richting geformuleerd. De Nederlandstalige items komen overeen met de gebruikelijke vertaling in het hedendaagse onderzoek (Scheepers & Eisinga, 1991, p. 190-196). De vertaling naar het Frans gebeurde aan de hand van de zogenaamde *committee approach*. De verschillende subsyndromen van het autoritarismesyndroom zijn duidelijk in de verkorte schaal vertegenwoordigd (Billiet, 1993, p.154). De respondenten dienden op een vijfpuntenschaal aan te geven in welke mate zij

het eens of oneens waren met de desbetreffende uitspraak, waarbij 1 stond voor 'helemaal eens' en 5 voor 'helemaal oneens'.

2.2. De onderzoekspopulatie

De hierboven besproken verkorte F-schaal werd voorgelegd aan een representatieve steekproef van de Belgische kiesgerechtigde burgers die in het kader van het verkiezingsonderzoek van 1991, naar aanleiding van de parlementsverkiezingen van 24 november van datzelfde jaar, werden bevraagd. Dit onderzoek werd uitgevoerd door het ISPO in samenwerking met zijn Franstalige tegenhanger PIOP. De steekproef bevat ongeveer 4 500 respondenten die werden geselecteerd op basis van de gegevens van het Rijksregister. Hiervoor werd gebruikgemaakt van een tweetrapssteekproef waarbij eerst op toevallige basis gemeenten werden geselecteerd en vervolgens werd, eveneens op toevalsbasis, bepaald welke inwoners van deze gemeenten zouden worden bevraagd. De steekproefrekking gebeurde zodanig dat elk van de elementaire steekproefeenheden een gelijke kans had om in de steekproef te worden opgenomen. Op die manier werden uiteindelijk 2 784 Nederlandstalige respondenten en 1 713 Franstalige respondenten bevraagd (2) (ISPO/PIOP, 1995) (3).

3. EVALUATIE VAN DE VERGELIJKBAARHEID VAN DE SCORES OP DE VERKORTE F-SCHAAL

Alvorens tot de analyse van de scores op de F-schaal over te gaan, werden enkele aanpassingen in de data doorgevoerd. Een eerste aanpassing betreft de respondenten die geen antwoord gaven of zeiden geen mening te hebben op minstens één van de zeven items van de F-schaal. Voor deze respondenten kan geen totale schaalscore worden berekend. Men kan dit probleem verhelpen door al deze respondenten gewoonweg uit de data te verwijderen. Dit zou echter leiden tot de verwijdering van 366 respondenten (ongeveer 8%) uit de dataset. Om het verlies van dit relatief groot aantal respondenten te vermijden, werd gebruikgemaakt van een imputatiemethode. Voor de respondenten met twee of minder ontbrekende waarden op de zeven items van de schaal en voor de respondenten met drie ontbrekende waarden en een variatiebreedte op de schaal die niet groter is dan drie, werden deze ontbrekende waarden vervangen door het afgerond gemiddelde van de antwoorden die de respondent gaf op de overige items van de F-schaal. Respondenten met meer dan drie ontbrekende waarden op de items van de F-schaal en respondenten met drie ontbrekende waarden en een variatiebreedte die groter is dan drie werden alsnog uit de dataset verwijderd. In zulke gevallen zou de imputatie op een te beperkt aantal items gebaseerd zijn of zou dit kunnen leiden tot inconsistente imputa-

ties. Bij respondenten met een relatief grote variatiebreedte kan men immers moeilijk voorspellen welk antwoord zij gegeven zouden hebben op het desbetreffende item met een ontbrekende waarde. Slechts 35 respondenten werden niet weerhouden. De uiteindelijke steekproef bevat dus 4 462 respondenten: 2 770 Nederlandstalige en 1 692 Franstalige respondenten. Bij 7.4% van deze respondenten werden de waarden voor één of meerdere items bekomen door de hierboven beschreven imputatiemethode. Op basis van deze steekproef zullen de analyses met betrekking tot de validiteit van de vergelijking van het autoritarismeniveau tussen Frans- en Nederlandstaligen worden gebaseerd. Ten slotte werden de items gehercodeerd zodanig dat een hogere score op een item ook verwijst naar een meer autoritaire houding. De waarde 1 betekent nu dus dat de respondent het 'helemaal oneens' was met de desbetreffende uitspraak. Score 5 betekent dat de respondent het 'helemaal eens' was met die uitspraak.

Laten we in eerste instantie de gemiddelde gesommeerde schaalscores van de Franstalige en de Nederlandstalige respondenten eens met elkaar vergelijken. Voor de Nederlandstalige respondenten komt men tot een gemiddelde score van 24.47 op de gesommeerde F-schaal. De gemiddelde score van de Franstalige respondenten ligt met een waarde van 25.34 significant hoger dan bij de Nederlandstalige respondenten ($T(3217.9) = -5.41; p < 0.0001$). Op basis van deze gegevens zou men dus kunnen besluiten dat de Franstalige respondenten een meer autoritaire houding hebben dan de Nederlandstalige respondenten. De vraag die we in deze bijdrage aan de orde stellen is echter of een dergelijke vergelijking wel is toegestaan. Indien men immers een dergelijke uitspraak doet, veronderstelt men dat de meetschalen in de beide onderzochte groepen op dezelfde ratioschaal worden uitgedrukt, met andere woorden dat geen enkele vorm van bias (mede)verantwoordelijk is voor het verschil tussen Frans- en Nederlandstalige respondenten.

In de volgende paragrafen gaan we na of de conclusie dat Franstaligen meer autoritair zijn dan Nederlandstaligen werkelijk gerechtvaardigd is. In eerste instantie zullen we nagaan of de betrouwbaarheid van de schalen in voldoende mate overeenstemt. Vervolgens zal worden ingegaan op de vergelijkbaarheid van de scores op beide schalen. Concreet betekent dit dat we zullen onderzoeken of de verschillen te wijten kunnen zijn aan *construct en/of item bias*.

3.1. De betrouwbaarheid van de schalen

"Cross-cultural comparisons of scores presuppose accurate psychometric properties of the measures in all groups. Preliminary analyses examine these properties at the item or instrument level. In the latter case, it is common to compare reliability coefficients" (van de Vijver & Leung, 1997, p.60). Zowel voor de Franstalige als de Nederlandstalige schaal werd de betrouwbaarheid berekend aan de hand van Cronbach's alpha (Billiet, 1990 (1994), p. 243-244). Voor de Nederlandstalige schaal bedraagt Cronbach's alpha 0.773 (α_1), voor de Franstalige schaal 0.791 (α_2). De

toetsingsgrootheid voor het beoordelen van de gelijkheid van twee onafhankelijke betrouwbaarheidscoëfficiënten is $(1-\alpha_1)/(1-\alpha_2)$. Bij grote steekproeven volgt deze toetsingsgrootheid een F-verdeling met N_1-1 en N_2-1 vrijheidsgraden, waarbij N_1 en N_2 verwijzen naar de respectieve steekproefomvang (van de Vijver & Leung, 1997, p. 60). Voor de beide betrouwbaarheidscoëfficiënten die hierboven werden berekend, is deze toetsingsgrootheid gelijk aan 1.086. De kans om een dergelijke waarde F-waarde te bekomen in twee groepen met deze omvang onder de nulhypothese is ongeveer gelijk aan 3% ($p=0.0301$). De betrouwbaarheidscoëfficiënten van de beide schalen verschillen dus significant van elkaar op een α -niveau van 0.05, maar niet op een α -niveau van 0.01. Vaak wordt bij grote steekproeven een strenger significantieniveau gehanteerd dan het gebruikelijke α -niveau van 0.05 omdat het discriminerend vermogen van een steekproef afneemt met de omvang. In grote steekproeven leiden zelfs triviale verschillen tot significante waarden voor de teststatistiek. Ook hier lijkt, gezien het geringe verschil tussen de waarden van beide betrouwbaarheidscoëfficiënten en het grote aantal respondenten, het hanteren van een strenger significantie criterium aangewezen. Omdat de coëfficiënten van beide schalen niet significant van elkaar verschillen op een α -niveau van 0.01 lijkt het ons niet onredelijk te besluiten dat deze coëfficiënten in voldoende mate aan elkaar gelijk zijn. Bovendien kan men stellen dat een verschil in betrouwbaarheid van 0.18 (0.791-0.773) praktisch gezien, weinig relevant is. Aan de eerste vereiste om de scores met elkaar te kunnen vergelijken is dus voldaan.

3.2. Construct bias

In deze paragraaf zal worden onderzocht of de Franstalige en de Nederlandstalige schaal voldoen aan het criterium van *construct equivalence*. We zullen met andere woorden nagaan of de Nederlandstalige en de Franstalige F-schaal eenzelfde latent concept meten. Deze evaluatie zal gebeuren aan de hand van twee methoden: factor-congruentieanalyse en de *linear structural relations (LISREL)* benadering. Een gedetailleerde bespreking van beide technieken vindt men bij Phalet (1993).

3.2.1. Factor-congruentieanalyse

Bij deze werkwijze wordt bij elk van de culturele groepen een exploratieve factoranalyse uitgevoerd met betrekking tot de schaal die men wil onderzoeken. De oplossingen van elk van de culturele groepen kunnen dan worden vergeleken aan de hand van verschillende congruentie-indices. Vaak manifesteert zich hierbij echter een *rotatieprobleem*. De ruimtelijke oriëntatie van de factoren is immers arbitrair. Om dit probleem te vermijden dienen de rotaties van de matrices met de factorladingen op elkaar te worden afgestemd. Dit gebeurt aan de hand van een zogenaamde *target rotation* of *procrustes rotation* (4) waarbij de factorladingen van de culturele groepen worden afgestemd op de factorladingen van één van de andere

culturele groepen of op een gemeenschappelijke matrix van factorladingen (van de Vijver & Leung, 1997, p. 90-99). In onze analyse stelt dit rotatieprobleem zich echter niet omdat in elk van de culturele groepen slechts één factor geanalyseerd dient te worden.

Uit deze exploratieve factoranalyse blijkt dat de items van de autoritarismeschaal zowel bij de Nederlandstalige respondenten als bij de Franstalige respondenten slechts door één achterliggend latent concept worden beïnvloed. De *scree plot* wijst duidelijk op de aanwezigheid van slechts één factor. Ook op basis van de proportie verklaarde variantie die de eerste factor op zich neemt en het eigenwaardecriterium kan men besluiten tot het weerhouden van slechts één factor. Bij de Franstalige respondenten is de eigenwaarde van de eerste factor gelijk aan 2.4163, de eigenwaarde van de tweede factor bedraagt slechts 0.0655. De eigenwaarde van de eerste factor bij de Nederlandstalige respondenten bedraagt 2.2712, die van de tweede factor is gelijk aan 0.1465. De factorladingen van de items op het latente concept voor beide groepen van respondenten vindt men in tabel 2.

Tabel 2. Factorladingen van de F-schaalitems op het latente concept

<i>Schaalitems</i>	<i>Nederlandstalige respondenten</i> (<i>N</i> =2 770)	<i>Franstalige respondenten</i> (<i>N</i> =1 692)
<i>Item 1</i>	0.66	0.67
<i>Item 2</i>	0.59	0.58
<i>Item 3</i>	0.47	0.50
<i>Item 4</i>	0.58	0.56
<i>Item 5</i>	0.56	0.63
<i>Item 6</i>	0.59	0.63
<i>Item 7</i>	0.52	0.53

Een algemeen aanvaard criterium voor het bepalen van een relevante factorlading bestaat niet. Op basis van de gebruikelijke onderzoekspraktijken stelt Waage (1996, p. 133) de waarde 0.45 voorop als minimumwaarde voor een relevante factorlading. Als strikter criterium hanteert deze auteur de waarde 0.60. Zoals men in de bovenstaande tabel kan vaststellen, zijn alle factorladingen groter dan 0.45. In beide groepen liggen vier van de zeven ladingen in de buurt van 0.60 of erboven. Alle items laden dus in voldoende mate op de achterliggende latente factor. Men kan tevens vaststellen dat de factorladingen in beide groepen vrij goed met elkaar overeenstemmen. Verschillende auteurs (Cavusgil, 1985; Ten Berge, 1986; van de Vijver & Leung, 1997; van de Vijver & Poortinga, 1994; Waage, 1996) beschrijven echter

een aantal meer formele criteria om de gelijkens tussen de factorladingen van cross-culturele groepen te evalueren.

Van de Vijver & Leung (1997) onderscheiden vier beschrijvende maten (of congruentie-indices) om de gelijkens tussen de factorladingen van de verschillende groepen met elkaar te vergelijken. Een eerste maat is de *identity coefficient*. Deze is gevoelig voor zowel additieve als voor multiplicatieve transformaties van de scores. Een volgende maat is de *additivity coefficient*. Deze is wel gevoelig voor multiplicatieve transformaties van de scores, maar niet voor additieve transformaties. De meest populaire maat voor een evaluatie van de gelijkens tussen de factorladingen is de *proportionality coefficient of Tucker's Phi*. *Tucker's coefficient* is niet gevoelig voor multiplicatieve transformaties in de factorladingen, echter wel voor additieve transformaties in de scores. Ten slotte beschrijven zij *Pearson's correlatiecoëfficiënt* als een maat voor de evaluatie van de gelijkens tussen de scores. Deze is ongevoelig voor zowel multiplicatieve als additieve transformaties van de scores. Om deze reden en omwille van het feit dat men negatieve deviaties van het gemiddelde bekommt indien hoge en kleine waarden met elkaar vergeleken worden, raadt Waegel (1996, p. 135) het gebruik van deze laatste coëfficiënt af. Van de Vijver & Leung (1997) erkennen de problemen die aan deze maten verbonden zijn en drukken er daarom op dat de evaluatie van de gelijkens tussen de scores dient te gebeuren door de verschillende coëfficiënten tegelijkertijd te evalueren. De waarden van deze maten voor de factorladingen op de autoritarismeschaal evenals de berekeningswijze van deze coëfficiënten worden weergegeven in tabel 3.

Tabel 3. Berekeningswijze en waarde van de maten van similariteit

Coëfficiënt	Berekeningswijze ^a	Waarde
<i>Identity</i>	$\frac{2 \sum x_i y_i}{\sum x_i^2 + \sum y_i^2}$	0.9983
<i>Additivity</i>	$\frac{2S_{xy}}{S_x^2 + S_y^2}$	0.8699
<i>Proportionality (Tucker's Phi)</i>	$\frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$	0.9988
<i>Linearity (Pearson's correlatie)</i>	$\frac{S_{xy}}{S_x S_y}$	0.8700

^a S_x = standaardafwijking van x; S_y = standaardafwijking van y;
 S_{xy} = covariantie van x en y

Een probleem bij deze coëfficiënten bestaat er echter in dat er geen duidelijke significantietest bestaat. Van de Vijver & Leung (1997, p. 92) stellen dat waarden boven de .95 voor deze coëfficiënten gezien kunnen worden als een indicatie voor factoriële gelijkens. Waarden onder .90 (van de Vijver & Poortinga, 1994) of onder .85 (Ten Berge, 1986) worden als problematisch beschouwd. Schneewind & Cattell (1970 in: Waage, 1996, p. 136) stellen echter dat een waarde van 0.78 voor *Tucker's Phi* overeenkomt met een significantie van .01 voor de similariteit van de schalen. Cavusgil (1985) stelt dat een waarde voor *Tucker's Phi* van 0.90 wijst op voldoende similariteit tussen de factorladingen. In tabel 3 kan men vaststellen dat de *identity coefficient* en *Tucker's Phi* bijna gelijk zijn aan 1. Beide andere coëfficiënten bereiken de waarde 0.95 niet, maar zijn wel groter dan de minimumgrens van 0.85 die werd vooropgesteld. We kunnen op basis van de factor-congruentieanalyses dus besluiten dat de factorladingen van de schalen voldoende gelijkens vertonen. Deze resultaten wijzen bijgevolg uit dat de verkorte autoritarismeschaal hetzelfde latente concept meet bij de Franstalige en de Nederlandstalige respondenten.

3.2.2. Linear structural relations benadering

Aan de werkwijze die in de vorige paragraaf werd gebruikt zijn evenwel enkele problemen verbonden. Ten eerste beschikt men bij exploratieve factoranalyse niet over formele criteria voor het testen van hypothesen omtrent het aantal te weerhouden factoren, de fit tussen het model en de data, enz. Een tweede probleem betreft de congruentie-indices die bij de vorige analyse werden gebruikt. Bijnen e.a. (1986) toonden immers aan dat ook op toevalsbasis hoge waarden voor deze indices bekomen kunnen worden en concluderen daarom dat: "*it is obvious that a lower limit of .95 for factor similarity is still not conservative enough in view of the high values found for data where any deviation of the similarity index from .00 obviously has no ... meaning*" (Bijnen e.a., 1986, p. 10). Beide problemen kunnen worden ondervangen met behulp van de *linear structural relations* benadering.

In deze analyse wordt eerst een model getest waarin de factorladingen en de varianties van de latente variabelen vrij worden geschat, met de restrictie dat de factorladingen in beide groepen aan elkaar gelijk worden gesteld. Deze analyse werd uitgevoerd met behulp van LISREL 8.12a. De schatting van de parameters gebeurde op basis van de covariantiematrix en de maximum likelihood schattingsprocedure. De parameterwaarde voor het eerste item werd voor elke groep vastgelegd op 1 zodanig dat de latente variabele in elke groep op dezelfde schaal wordt uitgedrukt (Jöreskog & Sörbom, 1989).

Tabel 4. Startmodel voor de gelijkheid van de factorladingen (Λ) in de Nederlandstalige en de Franstalige steekproef: niet-gestandaardiseerde parameters, t-waarden (tussen haakjes) en gestandaardiseerde parameters

Items	Λ : Nederlandstaligen (N=2 770)			Λ : Franstaligen (N=1 692)		
Item 1	1	(gefixeerd)	0.68	1	(gefixeerd)	0.68
Item 2	0.76	(32.45)	0.60	0.76	(32.45)	0.60
Item 3	0.66	(27.20)	0.49	0.66	(27.20)	0.49
Item 4	0.96	(31.87)	0.58	0.96	(31.87)	0.58
Item 5	0.92	(32.39)	0.60	0.92	(32.39)	0.60
Item 6	1.02	(33.03)	0.61	1.02	(33.03)	0.61
Item 7	0.81	(29.26)	0.53	0.81	(29.26)	0.53
Correlaties (ϕ)	0.46	(20.66)	1	0.62	(17.90)	1
Bijdrage Nederlandstaligen tot $\chi^2 = 272.74$ (75.58%)						
$\chi^2 = 360.88$, $df = 34$ ($p = 0.0$); RMSEA = .046, p-waarde voor close fit = .91; NNFI = .94; CFI = .95						

De resultaten van deze eerste analyse worden weergegeven in tabel 4. De waarden voor de NNFI (>.90) en de CFI (>.90) suggereren dat het model in voldoende mate aansluit bij de geobserveerde gegevens. De waarde voor de RMSEA is wel lager dan .05, maar de p-waarde voor de 'test of close fit' is eerder laag (deze p-waarde zou bijna gelijk moeten zijn aan 1) (5). De χ^2 -test wijst eveneens in de tegenovergestelde richting. Gezien de steekproefomvang is dit echter niet verwonderlijk. Vooral bij dit soort toepassingen is de χ^2 -statistiek, in tegenstelling tot de drie hierboven vernoemde maten, zeer gevoelig voor de steekproefomvang. Daarom streeft men er in de praktijk vaak naar om een model te bekomen met een χ^2 -waarde die relatief klein is ten opzichte van het aantal vrijheidsgraden. (Het probleem van de gevoeligheid voor de steekproefomvang blijft ook hier echter bestaan.) Als vuistregel wordt dan gesteld dat de verhouding tussen de χ^2 -waarde en het aantal vrijheidsgraden kleiner moet zijn dan twee (Hatcher, 1990, p.290). Voor het bovenstaande model is dit quotiënt echter beduidend groter dan de vooropgestelde waarde, wat opnieuw wijst op een te grote discrepantie tussen het gepostuleerde model en de data.

De modificatie-indices voor het bovenstaande model suggereerden echter dat de voornaamste oorzaak voor de relatief zwakke overeenstemming tussen het model en de data niet te wijten was aan de restrictie van gelijke factorladingen in de twee groepen, maar aan het feit dat dit eerste model geen gecorreleerde fouten toeliet. Indien gecorreleerde fouten tussen een aantal items van de F-schaal werden toegelaten, bleek het mogelijk te zijn om een model te vinden met een aanvaardbare waarde voor de χ^2 -toets en waarin de vereiste van gelijke factorladingen in de beide groepen behouden kon blijven. Voor dit model met gecorreleerde fouten (tabel 5) zijn de waarden van de NNFI, de CFI en de 'p-value of close fit' met betrekking tot de RMSEA (nagenoeg) maximaal. De verhouding tussen de waarde voor de χ^2 -test en het bijbehorende aantal vrijheidsgraden is kleiner dan 2. Zelfs de p-waarde van de χ^2 -statistiek is, rekening houdend met de steekproefgrootte (ongeveer 4 500 respondenten), aanvaardbaar te noemen. We kunnen met andere woorden op basis van dit model besluiten dat de Franstalige en de Nederlandstalige schalen een zelfde concept meten. Het resultaat van de factor-congruentieanalyse uit de vorige paragraaf wordt hier dus bevestigd.

Tabel 5. Meetmodel voor de gelijkheid van de factorladingen (Λ) in de Nederlandstalige en de Franstalige steekproef: niet-gestandaardiseerde parameters, t-waarden (tussen haakjes) en gestandaardiseerde parameters^a

Items	Λ : Nederlandstaligen (N=2 770)			Λ : Franstaligen (N=1 692)		
Item 1	1	(gefixeerd)	0.65	1	(gefixeerd)	0.65
Item 2	0.74	(31.61)	0.55	0.74	(31.61)	0.55
Item 3	0.70	(24.77)	0.49	0.70	(24.77)	0.49
Item 4	1.08	(25.94)	0.62	1.08	(25.94)	0.62
Item 5	1.11	(29.40)	0.68	1.11	(29.40)	0.68
Item 6	1.05	(25.48)	0.60	1.05	(25.48)	0.60
Item 7	0.81	(24.26)	0.50	0.81	(24.26)	0.50
Correlaties (ϕ)	0.42	(17.73)	1	0.55	(15.62)	1
Bijdrage Nederlandstaligen tot $\chi^2 = 24.52$ (60.79 %)						
$\chi^2 = 40.34$, $df = 22$ ($p = 0.01$); RMSEA = .014, p-waarde voor close fit = 1.00; NNFI = .99; CFI = 1.00						

^a de gecorreleerde fouten hebben betrekking op de itemparen (2,1), (6,4), (7,6), (5,1), (5,2) en (5,4).

Bij deze analyse dienen echter wel enkele bemerkingen geformuleerd te worden. Ten eerste stelt zich bij deze manier van werken het probleem van het herhaaldelijk testen. Het model wordt telkens op basis van modificatie-indices aangepast om uiteindelijk een model te bekomen dat in voldoende mate aansluit bij de geobserveerde data, waardoor de kans op een type I-fout over de verschillende testen heen gecumuleerd wordt. Een tweede bemerking geldt specifiek voor het hierboven geselecteerde model. In dit model kon de vereiste voor constructiequivalentie (gelijke factorladingen in beide culturele groepen) worden bekomen, door het toelaten van significant van nul verschillende foutencovarianties. Een dergelijke manier van werken is echter niet geheel onproblematisch. Gecorreleerde fouten tussen de items van een meetschaal wijzen er immers op dat deze items nog variantie gemeenschappelijk hebben die niet kan worden toegeschreven aan de achterliggende latente factor(en). Dit kan betekenen dat er nog een andere latente factor in de data onderkend zou kunnen worden dan degene die in het model zijn opgenomen. De exploratieve factoranalyse in de vorige paragraaf wees in beide groepen echter duidelijk op het bestaan van slechts één achterliggende factor. Bovendien blijken de meerderheid van de gecorreleerde fouten die werden toegestaan niet statistisch significant (vooral in de Franstalige steekproef), maar dienen ze wel behouden te blijven om een aanvaardbare fit te bekomen voor de χ^2 -statistiek. Opnieuw laat de grootte van de steekproef zich gevoelen.

3.3. Item bias

Item bias kan worden veroorzaakt door incidentele verschillen in de geschiktheid van de inhoud van een item of door een inadequate formulering of vertaling van een item. Er bestaan verschillende technieken om een dergelijke vorm van bias te onderzoeken. Voor een bespreking van deze technieken verwijzen we naar van de Vijver & Leung (1997, p. 62-88). Om na te gaan in welke mate de items van de verkorte autoritarismeschaal in de beide taalgroepen onderhevig zijn aan *item bias*, maken we hier gebruik van covariantieanalyses. Deze techniek werd verkozen boven een conditionele variantieanalyse omwille van de eerder arbitraire selectie van de scoregroepen in deze laatste analysetechniek. Voor de uitvoering van een conditionele variantieanalyse dient het scoreniveau op de gehele schaal immers ingedeeld te worden in groepen van scoreniveaus. Deze groepen kunnen echter op verschillende manieren worden gevormd: men kan de groepen indelen op basis van het aantal scoreniveaus maar ook op basis van het aantal respondenten. Daarnaast dient men nog te bepalen hoeveel groepen men zal weerhouden. De resultaten van een dergelijke analyse zijn daarom niet erg robuust: een andere groepsindeling kan gemakkelijk tot andere resultaten leiden. Om deze reden werd geopteerd voor covariantieanalyse.

In deze analyse worden telkens één afhankelijke, één onafhankelijke variabele en één covariaat betrokken. Eén van de items van de schaal zal fungeren als afhankelijke variabele. De onafhankelijke variabele heeft betrekking op de taalgroep waar-

toe de respondenten behoren. Het scoreniveau waartoe de respondent behoort, fungeert als covariaat in de analyse. Deze variabele werd geconstrueerd op basis van de som van de scores van de respondenten op alle items van de F-schaal, behalve het item dat in de desbetreffende analyse als afhankelijke variabele fungeert. Bij elk van de items van deze schaal konden de respondenten een antwoord geven aan de hand van een vijfpuntschaal. Indien de scores voor deze items worden opgeteld bekomt men de waarden 6 en 30 als minimum en maximum voor de gesommeerde schaal. Behalve voor deze minimale en maximale waarde kunnen deze gesommeerde scores echter op een zeer verschillende wijze totstandkomen. De score 21 kan men bijvoorbeeld bekomen indien de respondent telkens de waarde 3 als antwoord geeft, maar ook door bijvoorbeeld driemaal 5, tweemaal 2 en tweemaal 1 te antwoorden. Omwille van de invariantie in de totstandkoming van de minimum- en maximumwaarden van de totale schaalscores kan *item bias* daar niet worden nagegaan. Respondenten met een dergelijke score worden daarom uit de analyse geweerd.

Bij de covariantieanalyses die uitgevoerd zullen worden, fungeert een item van de F-schaal dus telkens als afhankelijke variabele. Het hoofdeffect van de covariaat (het scoreniveau) op de afhankelijke variabele zal waarschijnlijk steeds significant zijn: dit wijst er immers op dat degenen met een hoge score op de andere zes items van de autoritarismeschaal ook hoog scoren op het desbetreffende item. De twee andere effecten zijn relevanter voor onze analyse. De rationale achter deze werkwijze is immers dat personen met een zelfde score op het achterliggende concept, een zelfde score dienen te hebben op elk item van de *measurement scale* ongeacht de culturele groep waartoe ze behoren. Indien zowel het hoofdeffect van de taalgroep waartoe de respondent behoort als het interactie-effect tussen de taalgroep en het scoreniveau niet significant zijn dan is het item onvertekend. Indien enkel het hoofdeffect van de taalgroep waartoe de respondent behoort significant is, dan heeft men te maken met een uniforme vertekening. Dit betekent immers dat respondenten met een zelfde totale schaalscore, hoger of lager scoren op dat item naargelang van de taalgroep waartoe ze behoren. Is het interactie-effect tussen de taalgroep en de scoregroep waartoe de respondenten behoren statistisch significant, dan heeft men te maken met een niet-uniforme vertekening. De relatie tussen de score op een bepaald item en het scoreniveau op de rest van de schaal verschilt dan immers naargelang van de taalgroep waartoe de respondent behoort.

Eén van de assumpties in een covariantieanalyse is dat de regressielijnen die de relatie tussen de covariaat en de (metrische) afhankelijke variabele uitdrukken in elke van de culturele groepen, parallel aan elkaar zijn. Deze assumptie veronderstelt met andere woorden dat er geen interactie is tussen de covariaat en de onafhankelijke variabele (Anderson e.a., 1980; Kirk, 1968 (1995)). Indien een dergelijke interactie optreedt, is er geen unieke samenvattende waarde die geïnterpreteerd kan worden als het hoofdeffect van het behoren tot deze of gene culturele groep (de onafhankelijke variabele). Voor onze analyses is dit problematisch omdat we zowel geïnteresseerd zijn in *het* hoofdeffect van de onafhankelijke variabele als het

interactie-effect tussen de onafhankelijke variabele en de covariaat. In dergelijke gevallen dient het hoofdeffect van de culturele groep op de score op een bepaald item nader bekeken te worden.

Tabel 6. Covariantie-analyses per item

Item	N	Hoofdeffect taalgroep			Interactie taalgroep/scoreniveau		
		F-waarde	p-waarde	R ² ^d	F-waarde	p-waarde	R ² ^d
Item 1 ^a	4 372	16.50	0.0001	0.0080	6.86	0.0089	0.0033
Item 2 ^b	4 355	0.34	0.5596	0.0003	0.49	0.4830	0.0004
Item 3 ^b	4 362	2.43	0.1193	0.0028	1.52	0.2184	0.0017
Item 4 ^b	4 355	0.18	0.6724	0.0001	0.12	0.7277	0.0001
Item 5 ^c	4 361	4.00	0.0454	0.0025	8.55	0.0035	0.0053
Item 6 ^b	4 359	0.18	0.6726	0.0001	2.14	0.1433	0.0012
Item 7 ^b	4 323	4.29	0.0385	0.0040	0.87	0.3504	0.0008

^a beïnvloed door een uniforme en een niet-uniforme vertekening.

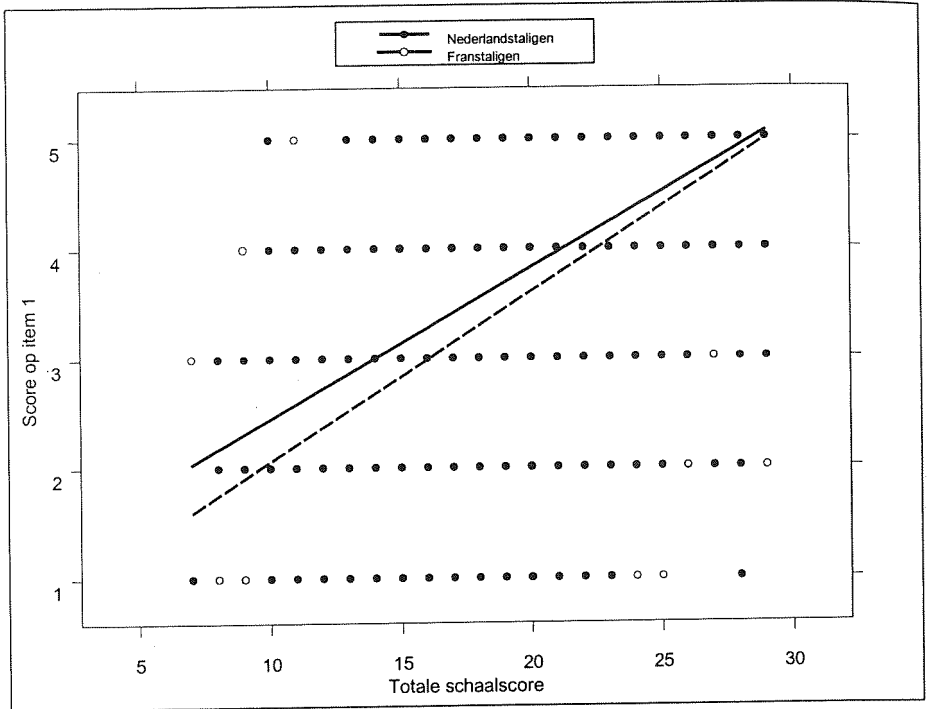
^b onvertkend.

^c beïnvloed door een niet-uniforme vertekening.

^d kwadratensom van het effect gedeeld door de kwadratensom van het model (aandeel van het effect in de verklaarde variantie).

De resultaten van de covariantieanalyses voor de items van de autoritarismeschaal, worden weergegeven in tabel 6. Omwille van de grootte van de steekproef werd opnieuw een significantieniveau van 0.01 gehanteerd. De resultaten tonen aan dat de items 2, 3, 4 en 6 niet vertekend worden door item bias. Bij deze items is zowel het hoofdeffect van de taalgroep als het interactie-effect tussen de taalgroep en het scoreniveau waartoe de respondent behoort, niet statistisch significant. Bij het eerste item is zowel de interactie tussen de covariaat en de onafhankelijke variabele als het hoofdeffect van de onafhankelijke variabele statistisch significant op een α -niveau van 0.01. Dit item wordt dus zowel door een uniforme als door een niet-uniforme bias vertekend. Bij het vijfde item is enkel de interactie tussen de onafhankelijke variabele en de covariaat statistisch significant op een α -niveau van 0.01. Dit item wordt met andere woorden enkel beïnvloed door een niet-uniforme vertekening. Als we echter nagaan welk aandeel van de verklaarde variantie in het respectieve model op rekening kan worden geschreven van de methode-effecten, blijkt dat de impact van deze methode-effecten vrij gering is. Geen enkel van de significante effecten neemt meer dan één procent van de verklaarde variantie voor zijn rekening.

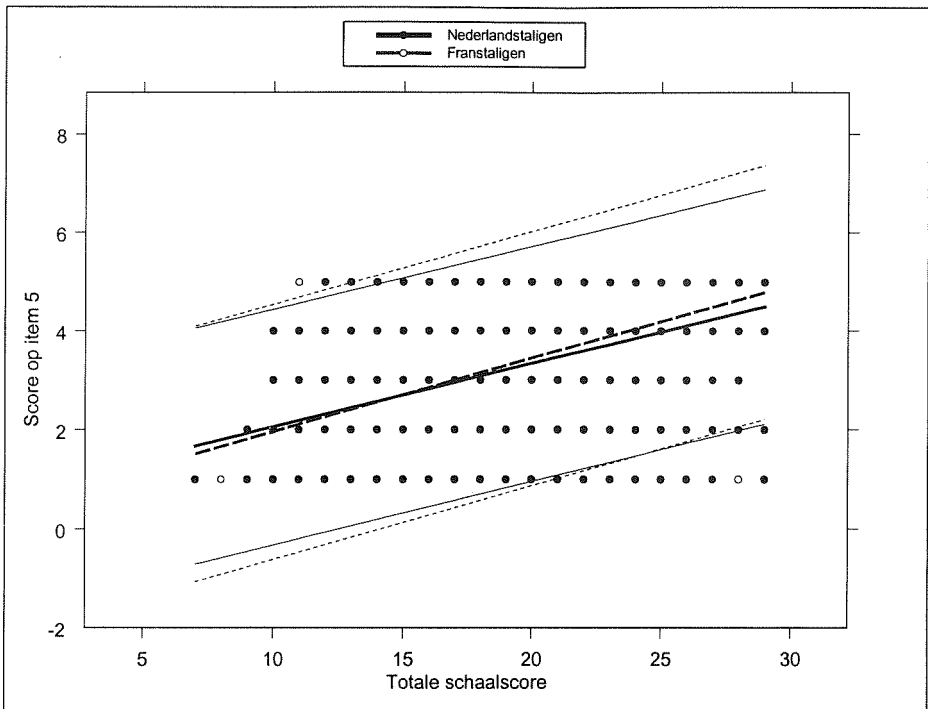
Figuur 1. De relatie tussen de score op item 1 en de covariaat naargelang van de taalgroep



Zoals reeds gesteld dienen de hoofdeffecten van de onafhankelijke variabele nader bekeken te worden indien ook het interactie-effect tussen de covariaat en de onafhankelijke variabele statistisch significant is. Zolang de interactie tussen covariaat en de onafhankelijke variabele niet de vorm aanneemt van een zogenaamde kruisinteractie stellen er zich weinig problemen bij de interpretatie van het hoofdeffect van de onafhankelijke variabele. In dat geval liggen de scores van de ene culturele groep op de afhankelijke variabele steeds hoger of lager dan de scores van de andere groep, ongeacht de waarde van de covariaat. Zoals blijkt uit figuur 1 treedt er bij de covariantieanalyse met betrekking tot het eerste item geen kruisinteractie op: de scores van de Nederlandstalige respondenten op dit item liggen steeds hoger dan de scores van de Franstalige respondenten, ongeacht het scoreniveau. Bij het vijfde item treedt er echter wel een kruisinteractie op (figuur 2). Om na te gaan voor welke waarden van de totale schaalscore er een significant verschil is tussen de Franstalige en de Nederlandstalige respondenten, werden rond de regressielijnen van beide groepen betrouwbaarheidslijnen geconstrueerd die overeenkomen met een α -niveau van .01. Omwille van de kleinere omvang van de Franstalige steekproef zijn de betrouwbaarheidsintervallen rond de regressielijn van deze groep

steeds groter dan die voor de Nederlandstalige respondenten. Deze betrouwbaarheidslijnen suggereren dat de Franstalige respondenten met een totale schaalscore van ongeveer 24 en meer significant hoger scoren op het vijfde item dan de Nederlandstalige respondenten met een zelfde totale schaalscore. Bij de betrouwbaarheidslijnen die als bovengrens fungeren voor de respectieve regressielijnen ligt de betrouwbaarheidslijn van de Franstaligen steeds hoger dan deze van de Nederlandstalige respondenten. Bij de betrouwbaarheidslijnen die fungeren als ondergrens voor de regressielijnen is dit echter slechts het geval vanaf ongeveer de waarde 24. Bij de hogere schaalscores lijkt er met andere woorden wel een significant hoofdefect te bestaan van de taalgroep waartoe de respondenten behoren.

Figuur 2. De relatie tussen de score op item 5 en de covariaat naargelang van de taalgroep



De vraag stelt zich nu uiteraard waardoor de vertekening bij het eerste en het vijfde item kan zijn teweeggebracht. Een nadere inspectie van de Nederlandstalige en de Franstalige versies van de items wijst in eerste instantie op een discrepantie wat betreft de antwoordmogelijkheden. Bij de Nederlandstalige versie konden de respondenten antwoorden met 'helemaal eens', '*eens*', 'noch eens, noch oneens', 'oneens', 'helemaal oneens'. De Franstalige respondenten konden antwoorden met 'tout à fait d'accord', '*plutôt d'accord*', 'ni d'accord, ni en désaccord', 'pas d'accord',

'pas du tout d'accord'. Aan de bevestigende zijde komen de Nederlandstalige en de Franstalige antwoordschalen dus niet met elkaar overeen. Waar in de Nederlandse versie de term 'eens' werd gebruikt, was in de Franse versie de antwoordmogelijkheid 'eerder eens'. Aan de andere zijde van de schaal komen de antwoordmogelijkheden wel met elkaar overeen. Deze discrepantie aan de bevestigende zijde van de schaal zou verantwoordelijk kunnen zijn voor de hierboven vastgestelde niet-uniforme item bias bij beide items. Een dergelijke verklaring voor de gevonden niet-uniforme vertekening bij deze items, lijkt echter onwaarschijnlijk omdat deze discrepantie in de antwoordschaal optreedt bij alle items, terwijl enkel bij het eerste en het vijfde item een niet-uniforme vertekening werd vastgesteld. Het zou echter ook mogelijk kunnen zijn dat juist bij deze twee items meer gebruik werd gemaakt van de bevestigende zijde van de antwoordschaal dan bij de andere vijf items. In dat geval zou de discrepantie in de Franstalige en Nederlandstalige antwoordschalen wel een geldige verklaring kunnen zijn voor de niet-uniforme vertekening die bij item 1 en 5 werd vastgesteld. Niets wijst er echter op dat dit het geval is: op het eerste en het vijfde item werd over het algemeen niet vaker bevestigend geantwoord dan op de vijf overige items. 72.3 en respectievelijk 58.7 procent van de respondenten antwoorden met 'helemaal eens' of '(eerder) eens' op het eerste en het vijfde item. Bij item 2,3,4,6 en 7 antwoordde respectievelijk 69.9, 73.1, 53.1, 46.8 en 35.7 procent van de respondenten dat ze het 'helemaal eens' of '(eerder) eens' waren met de gegeven uitspraak.

Een verklaring voor de uniforme bias die optreedt bij het eerste item van de schaal is evenmin voor de hand liggend. Waarschijnlijk zijn subtiele betekenisverschillen in de formulering van dit item de oorzaak voor de uniforme bias die aldaar werd vastgesteld. Zo werd de term 'het gezag' die in de Nederlandse versie verschijnt, naar het Frans vertaald door de term 'l'autorité'. Het is mogelijk dat de connotatie van de Nederlandse term 'gezag' net iets anders is dan de connotatie van de term 'l'autorité'. Naast de vertalingsprocedures zoals de *translation-backtranslation* procedure en de *committee approach* (die voor de vertaling van de hier gebruikte F-schaal werd toegepast), zouden zogenaamde cognitieve interviews een inzicht kunnen bieden in de subtiele betekenisverschillen van bepaalde termen (6).

4. CONCLUSIES

In deze bijdrage werd nagegaan in welke mate scores van Nederlandstalige en Franstalige respondenten op een verkorte F-schaal, die hen werd voorgelegd in het kader van het verkiezingsonderzoek van 1991, met elkaar vergelijkbaar zijn. Concreet werd onderzocht of de scores van de twee taalgroepen beïnvloed werden door *construct* en/of *item bias*. In eerste instantie diende echter nog te worden nagegaan of de betrouwbaarheid van de Nederlandstalige en de Franstalige schaal vergelijk-

baar was. Aan deze vereiste werd voldaan: de betrouwbaarheid van de schalen verschilde niet significant voor de beide taalgroepen.

Om na te gaan of de Franstalige en de Nederlandstalige schalen hetzelfde concept meten werden twee technieken toegepast. Deze wezen beide in de richting van een voldoende gelijkenis tussen de factorladingen van beide culturele groepen. Bij de LISREL-analyse kon een (volgens de χ^2 -statistiek) aanvaardbaar model echter pas bekomen worden op voorwaarde dat het optreden van gecorreleerde fouten tussen een aantal items werd toegelaten. Verschillende auteurs (McCrae e.a., 1996; Katigbak e.a., 1996) hebben trouwens reeds geargumenteed dat voor de evaluatie van de constructequevalentie van een meetschaal, een factor-congruentieanalyse met procrustes rotatie (7) betere resultaten oplevert dan een confirmatorische factoranalyse. De restricties die deze laatste techniek oplegt, zouden te stringent en vaak onrealistisch zijn.

Voor het opsporen van vertekeningen op het niveau van de individuele items werd gebruikgemaakt van covariantieanalyses. Deze stelden een significante uniforme en niet-uniforme vertekening vast bij het eerste item en een significante niet-uniforme vertekening bij het vijfde item. Dit zijn tevens de items waarvoor in de LISREL-analyse het optreden van gecorreleerde fouten diende te worden toegelaten. In die zin is er dus wel enige overeenstemming tussen de beide technieken. Het aandeel van deze methode-effecten in de verklaarde variantie is echter zeer gering en bijgevolg weinig betekenisvol. Vandaar dat men ons inziens, onder voorbehoud van het optreden van method bias, kan concluderen dat de scores van de Franstalige en Nederlandstalige respondenten op de verkorte autoritarismeschaal rechtstreeks met elkaar vergeleken kunnen worden. De reeds hoger geopperde conclusie dat de Franstalige respondenten een meer autoritaire houding hebben dan de Nederlandstalige respondenten lijkt, opnieuw onder voorbehoud van het optreden van method bias, met andere woorden gegrond.

VOETNOTEN

- (1) Met dank aan Dr. H. Waage, Dr. G. Loosveldt, Dr. J. Billiet en de anonieme referees van het tijdschrift voor hun waardevolle bedenkingen.
- (2) De veertien respondenten bij wie de vragenlijst in het Duits werd afgenomen, werden niet in de analyse betrokken.
- (3) De data die in deze publicatie werden gebruikt werden beschikbaar gesteld door het ISPO (Interuniversitair Steunpunt voor Politieke opinie Onderzoek), dat wordt gesponsord door de federale diensten voor technische, culturele en wetenschappelijke aangelegenheden. De originele data werden verzameld door Jaak Billiet, Marc Swyngedouw, Ann Carton en Roeland Beerten. Noch

deze personen, noch het ISPO dragen enige verantwoordelijkheid voor de analyses en de interpretaties die hier worden gepresenteerd.

- (4) Een voorbeeld van een dergelijke *target rotation* met behulp van de SAS[®]/IML module vindt men bij McCrae e.a. (1996, p. 566 (appendix)).
- (5) De comparative fit index (CFI) en de non normed fit index (NNFI) zijn beide indices die uitdrukken hoeveel beter de fit van het geschatte model is dan de fit van een bepaald basismodel, meestal het onafhankelijkheidsmodel (dit is het model waarin alle geobserveerde variabelen ongecorrleerd zijn). Voor beide worden waarden boven .90 beschouwd als aanduiding voor een aanvaardbare fit. De Root mean square error of approximation (RMSEA) is een maat voor de discrepantie tussen het model en de populatie uitgedrukt per vrijheidsgraad. Waarden onder .05 worden beschouwd als een aanduiding van een aanvaardbare fit. De 'p-value of close fit' geeft de kans weer dat de RMSEA kleiner is dan .05. (Jöreskog & Sörbom, 1993)
- (6) Voor een voorbeeld van de toepassingen van cognitieve interviewtechnieken verwijzen we naar Nuyts e.a. (1997).
- (7) Een procrustes rotatie was voor onze analyse niet nodig omdat er slechts een latente factor in de analyse betrokken was.

BIBLIOGRAFIE

- ADORNO, T.W., E. FRENKEL-BRUNSWIK, D.J. LEVINSON & R.N. SANFORD (1950 (1982)), *The authoritarian personality. Abridged Edition*. New York: W.W. Norton & Company.
- ANDERSON, S., A. AUQUIER, W.W. HAUCK, D. OAKES, W. VANDAELE & H. WEISBERG (1980), *Statistical methods for comparative studies. Techniques for bias reduction*. New York: John Wiley & Sons.
- BERRY, J.W., Y.H. POORTINGA, M.H. SEGALL & P.R. DASEN (1992 (1993)), *Cross-cultural psychology. Research and applications*. Cambridge: Cambridge University Press.
- BIJNEN, E.J., T.Z.J. VAN DER NET & Y.H. POORTINGA (1986), 'On cross cultural comparative studies with the Eysenck Personality Questionnaire', *The Journal of Cross-Cultural Psychology*, 17: 3-16.
- BILLIET, J. (1990 (1994)), *Methoden van sociaal-wetenschappelijk onderzoek: ontwerp en dataverzameling*. Leuven: ACCO.
- BILLIET, J. (1993), 'Stabiliteit en verandering in de attitude tegenover 'vreemdelingen'', pp. 147-162, in: SWYNGEDOUW, M., J. BILLIET, A. CARTON, R. BEERTEN (eds.), *Kiezen is verliezen*. Leuven: ACCO.
- BILLIET, J. & H. WAEGE (1991 (1992)), *Het SAS[®]-systeem. Gegevensverwerking met de computer in de humane wetenschappen*. Leuven: ACCO.

- CARTON, A., M. SWYNGEDOUW, J. BILLIET & R. BEERTEN (1993), *Source Book of the Voters' Study in connection with the 1991 General Election*. Leuven: Socio-logisch Onderzoeksinstituut K.U.Leuven/ISPO.
- CAVUSGIL, S.T. (1985), 'Factor congruency analysis: a methodology for cross-cultural research', *Journal of the Market research Society*, 27(2): 147-155.
- HATCHER, L. (1994), *A step-by-step approach to using the SAS® system for factor analysis and structural equation modeling*. SAS® Institute Inc.
- ISPO/PIOP (1995), *1991 General Election Study Belgium. Codebook and Questionnaire*. Leuven: ISPO/PIOP.
- JOHNSON, T.P. (1998), 'Approaches to equivalence in cross-cultural and cross-national survey-research', *ZUMA Nachrichten Spezial, Cross-cultural survey equivalence*, (3): 1-40.
- JÖRESKOG, K.G. & D. SÖRBOM (1989), *LISREL® 7 user's reference guide*. Chicago: Scientific Software International Inc.
- JÖRESKOG, K.G. & D. SÖRBOM (1993), *LISREL® 8: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International Inc.
- KATIGBAK, M.S., A.T. CHURCH & T.X. AKAMINE (1996), 'Cross-cultural generalizability of personality dimensions: relating indigenous and imported dimensions in two cultures', *Journal of Personality and Social Psychology*, 70(1): 99-114.
- KIRK, R.E. (1968 (1995)), *Experimental design. Procedures for the behavioral sciences*. Pacific Grove: Brooks/Cole Publishing Company.
- LITTLE, T.D. (1997), 'Mean and covariance structures (MACS) analyses of cross-cultural data: practical and theoretical issues', *Multivariate Behavioral Research*, 32(1): 53-76.
- MCCRAE, R.R., A.B. ZONDERMAN, P.T. COSTA JR., M.H. BOND & S.V. PAUNONEN (1996), 'Evaluating replicability of factors in the revised NEO personality inventory: confirmatory factor analysis versus procrustes rotation', *Journal of Personality and Social Psychology*, 70(3): 552-566.
- MELOEN, J.D. (1983), *De autoritaire reactie in tijden van welvaart en crisis*. Amsterdam: Universiteit van Amsterdam, Dissertatie.
- NUYTS, K., H. WAEGE, G. LOOSVELDT & J. BILLIET (1997), 'Het gebruik van cognitieve interview-technieken bij het ontwikkelen en testen van meetinstrumenten voor survey-onderzoek', *Tijdschrift voor Sociologie*, 18(4): 477-500.
- PHALET, K. (1983), *Culturele waarden en persoonlijke keuzes: groepsloyaliteit en prestatiemotivatie bij Turkse en Belgische Jongeren*. Leuven: ongepubliceerd doctoraatsproefschrift.
- POORTINGA, Y.H. (1989), 'Equivalence of cross-cultural data: an overview of basic issues', *International Journal of Psychology*, 24(6): 737-756.
- SAS® Institute Inc. (1990 (1994)), *SAS/STAT® User's Guide, Version 6, Fourth Edition*. Cary: SAS® Institute Inc.
- SCHEEPERS, P. & R. EISINGA (eds.) (1991), *Onderdanig en intolerant. Lacunes en controverses in autoritarisme-studies*. Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen.

- SCHNEEWIND, K. & R.B. CATTELL (1970), 'Zum Problem der Factoridentification: Verteilungen und Vertranensintervalle von Kongruenzkoeffizienten', *Psychologische Beiträge*, 12: 214-226.
- SWYNGEDOUW, M., J. BILLIET, A. CARTON & R. BEERTEN (eds.) (1993), *Kiezen is verliezen*. Leuven: ACCO.
- TEN BERGE, J.M.F. (1986), 'Rotatie naar perfecte congruentie en de multipele groep methode', *Nederlands Tijdschrift voor de Psychologie*, 41: 218-225.
- VAN DE VIJVER, F.J.R. (1998), 'Towards a theory of bias and equivalence', *ZUMA Nachrichten Spezial, Cross-cultural survey equivalence*, (3): 41-65.
- VAN DE VIJVER, F. & K. LEUNG (1997), *Methods and Data Analysis for Cross-Cultural Research*. Thousand Oaks: Sage.
- VAN DE VIJVER, F. & Y. POORTINGA (1994), 'Methodological issues in cross-cultural studies on parental rearing behavior and psychopathology', p. 173-197, in: PERRIS, C., A. ARRINDELL & M. EISEMANN (eds.), *Parental rearing and psychopathology*. Chicester: John Wiley.
- VAN DIEST, S. (1993), *Enquête nationale sur les opinions, les attitudes et les comportements politiques et sociaux des Belges. Codebook Wallonie*. Louvain-La-Neuve: PIOP.
- WAEGE, H. (1996), 'The validity of unidimensional scales. Different association structures in relevant subgroups of a heterogeneous population', pp. 127-151, in: FERLIGOJ, A. & A. KRAMBERGER (eds.). *Developments in data analysis*. Ljubljana: FDV.
- WELKENHUYSEN-GYBELS, J. (1997), 'Van de Vijver, F. & Leung, K., *Methods and Data Analysis for Cross-Cultural Research*. Thousand Oaks: Sage, 200 p., boekbespreking', in: *Tijdschrift voor Sociologie*, 18(4): 523-525.

BIJLAGE

Nederlandstalige versie

Tussen haakjes de nummering zoals bij Adorno e.a. (Scheepers & Eisinga, 1991, p. 190-196).

En wat is uw persoonlijke mening over de volgende uitspraken?

U kan terug kaart nr. 29 gebruiken om te antwoorden.

1. Gehoorzaamheid en respect voor het gezag zijn de twee belangrijkste deugden die kinderen moeten leren. (1)
2. Jonge mensen zijn vaak opstandig maar als zij ouder worden moeten zij zich aanpassen. (21)
3. Wat we nodig hebben zijn minder wetten en instellingen en meer moedige en toegewijde leiders, waar het volk vertrouwen in kan hebben. (23)
4. Er zijn twee soorten mensen: sterken en zwakkelingen. (26)
5. Onze sociale problemen zouden grotendeels zijn opgelost, als we ons op de één of andere manier konden ontdoen van immorele en oneerlijke mensen. (34)
6. Als de mensen minder zouden praten en harder zouden werken, zou alles beter gaan. (37)
7. De meeste mensen vallen tegen als je ze beter leert kennen. (43) *Dit item werd enigszins aangepast omdat de originele versie 'Familiarity breeds contempt' zich moeilijk laat vertalen.*

Kaart 29

1. Helemaal eens
2. Eens
3. Noch eens, noch oneens
4. Oneens
5. Helemaal oneens

Franstalige versie

Tussen haakjes de nummering zoals bij Adorno e.a. (Scheepers & Eisinga, 1991, p. 190-196).

Et quelle est votre opinion personnelle en ce qui concerne les affirmations suivantes? Vous pouvez encore utiliser la carte n° 29 pour répondre.

1. L'obéissance et le respect de l'autorité sont les deux vertus les plus importantes que les enfants doivent apprendre. (1)
2. Les jeunes sont souvent révoltés mais ils doivent bien s'adapter quand ils vieillissent. (21)
3. Mieux vaut de bons dirigeants honnêtes et courageux que des lois et des institutions. (23)
4. Il y a deux types d'hommes: les forts et les faibles. (26)
5. Nos problèmes sociaux pourraient être résolus si d'une manière ou d'une autre on pouvait se débarrasser des gens malhonnêtes et sans moralité. (34)
6. Si les gens travaillaient plus et parlaient moins, tout irait mieux. (37)
7. La plupart des gens sont décevants quand on les connaît un peu mieux. (43) *Dit item werd enigszins aangepast omdat de originele versie 'Familiarity breeds contempt' zich moeilijk laat vertalen.*

Kaart 29

1. Tout à fait d'accord
2. Plutôt d'accord
3. Ni d'accord ni en désaccord
4. Pas d'accord
5. Pas du tout d'accord