Een evaluatie van de crossculturele equivalentie van meetschalen rekening houdend met methode-effecten

Jerry Welkenhuysen-Gybels en Jaak Billiet[1]

Summary

An evaluation of the cross-cultural equivalence of measurement scales taking method effects into account

Often it is assumed that scores from different cultural groups and/or language groups on a certain measurement scale or a certain item can be compared in a direct and straightforward way. However, this is not necessarily so. After all, such a direct comparison presupposes a certain degree of equivalence of the measurement scale or the item in question. This article evaluates to what extent scores of Flemish and Walloon respondents on a certain attitude scale are comparable. We investigate whether the measurement scale measures the same concept in both groups. We also account for a certain method effect (acquiescence), which allows us to evaluate the cross-cultural equivalence of the constructs more adequately and to find out whether respondents from both groups are equally prone to this type of method effect. Finally, an item bias evaluation is performed on the individual items of the measurement scale. These analyses lead us to conclude that it is justified to compare scores of Flemish and Walloons in a direct manner.

1. Inleiding

In de internationale literatuur besteedt men steeds meer aandacht aan het crosscultureel onderzoek en aan de (methodologische) problemen die daaraan verbonden zijn. Dit soort studies wordt immers, 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 (Van de Vijver & Leung, 1997; Welkenhuysen-Gybels, 1997).

Het basisprobleem van het cultuurvergelijkend onderzoek heeft betrekking op de vraag of crosscultureel onderzoek überhaupt wel mogelijk is. De meningen daarover zijn uiteraard verdeeld en zijn te situeren op een continuüm met aan de ene kant de idiografische en aan de andere kant de nomothetische opvatting (Van de Vijver & Hutschemaekers, 1990). Volgens de idiografische opvatting hebben concepten en methoden geen universele geldigheid, maar

zijn ze enkel bruikbaar binnen een bepaalde culturele setting. De nomothetische opvatting daarentegen verdedigt de universaliteit van concepten en methoden; een opvatting afkomstig van de natuurwetenschappen. In deze traditie wordt de universele geschiktheid van concepten en methoden als vanzelfsprekend aangenomen. Verschillende auteurs nemen echter een tussenliggende positie in (zie onder meer: Van de Vijver & Leung, 1997; Welkenhuysen-Gybels, 1998) en gaan uit van het standpunt dat crossculturele vergelijkingen mogelijk zijn op voorwaarde dat een voldoende mate van equivalentie tussen de concepten kan worden aangetoond.

Het begrip equivalentie wordt echter op verschillende manieren gedefinieerd (Johnson, 1998). Van de Vijver en Leung (1997) 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 een zelfde intervalschaal kan worden uitgedrukt (de afstand tussen de meeteenheden van de meetschaal moet dus hetzelfde zijn in elke culturele groep). Dit equivalentieniveau impliceert niet dat de meetschalen in elke groep een zelfde 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.* 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.

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). Een eerste vorm van vertekening is *construct bias.* Deze doet zich onder meer voor wanneer het construct niet identiek is voor alle culturele groepen of indien 'brede' concepten aan de hand van een (te) beperkt aantal items worden bevraagd. *Method bias* vloeit voort uit verschillen in de steekproeftrekking, de specifieke eigenschappen van het instrument of de afname ervan. Verschillen in antwoordstijlen (tussen culturele groepen), zoals het geven van sociaal wenselijke antwoorden of het gebruik van de extremen van een schaal, kunnen eveneens aanleiding geven 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 vertekening. 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 naar gelang het niveau van de scores.

2. Probleemstelling

Deze bijdrage zal nagaan in welke mate de scores van twee culturele groepen, Vlamingen en Walen, op een attitudeschaal met betrekking tot de houding ten opzichte van migranten met elkaar vergeleken kunnen worden. Meet men met de gebruikte meetschaal wel hetzelfde in beide groepen? Zijn uitspraken van het type 'Vlamingen voelen zich meer/minder bedreigd door migranten dan Walen' gerechtvaardigd op basis van de beschikbare data? We zullen met andere woorden illustreren hoe de 'gulden middenweg' tussen de idiografische en de nomothetische traditie kan bewandeld worden. Hierbij zal het kader dat door Van de Vijver en Leung (1997) werd aangereikt, gebruikt worden om een zo volledig mogelijke toetsing van de equivalentie van de meetschalen uit te voeren. Daarom zal in eerste instantie worden nagegaan of de constructen die door de schalen gemeten worden voldoende gelijkend zijn in beide culturele groepen. Deze toetsing zal gebeuren aan de hand van *structural equation modelling*. Het testen van construct-equivalentie, wat in de praktijk neerkomt op het testen op de afwezigheid van construct bias, kan via verschillende technieken gebeuren (zie onder meer: Van de Vijver & Leung, 1997; Welkenhuysen-Gybels, 1998). Het gebruik van structural equation modelling laat ons echter niet enkel toe om de construct-equivalentie van de gebruikte schalen te testen. Deze techniek laat in zekere mate ook toe om te controleren voor method bias. Survey onderzoek maakt immers vaak gebruik van reeksen van attitude-items die een zelfde antwoordschaal hanteren. Deze antwoordschalen hebben dan vaak een Likert-formaat, waarbij de respondenten dienen aan te geven in welke mate zij het eens of oneens zijn met de desbetreffende uitspraak. Verschillende (sociologische) auteurs hebben geargumenteerd dat dit soort items vatbaar is voor *acquiescence*, of de zogenaamde tendens tot instemmen. Deze antwoordstijl bestaat in de tendens tot het beamen van uitspraken of vragen, ongeacht de inhoud van deze vragen of uitspraken. Billiet en McClendon (1998, 1999) toonden aan dat het via structurele modellen mogelijk is om dit soort van antwoordstijl te identificeren en bijgevolg om ervoor te corrigeren. De identificatie van acquiescence via structural equation modelling vereist wel dat de gebruikte attitudeschalen gebalanceerd zijn. Dit wil zeggen dat de schalen ongeveer evenveel positief als negatief verwoorde items dienen te bevatten. Respondenten die een neiging hebben tot 'ja-zeggen' zullen dan zowel instemmen met de positief verwoorde items als met de negatief verwoorde items. Door het controleren voor deze vorm van method bias, kan de construct-equivalentie van de latente concepten tevens op een meer valide manier geëvalueerd worden.

Een rechtstreekse vergelijking van de scores van respondenten uit de beide groepen op de attitudeschalen is echter slechts toegestaan indien ook de individuele items in voldoende mate aan elkaar equivalent zijn. Net als bij de toetsing van construct-equivalentie, wordt daarbij omgekeerd tewerkgegaan: men gaat na of de items worden beïnvloed door één van de vormen van vertekening die leiden tot item-inequivalentie. Ook hier kan men op verschillende manieren tewerkgaan. In deze bijdrage maken we gebruik van covariantie analyses.

3. Data

Voor de analyses werd gebruik gemaakt van de gegevens die in het kader van het verkiezingsonderzoek naar aanleiding van de Belgische parlementsverkiezingen van 1995 werden verzameld. De steekproeven in Vlaanderen en Wallonië bestaan gedeeltelijk uit panelrespondenten van het verkiezingsonderzoek van 1991 en gedeeltelijk uit een nieuwe steekproef, zodat ook de nieuwe kiezers vertegenwoordigd zijn. De steekproeven zijn toevallig getrokken uit het Rijksregister en zijn representatief voor de Belgische kiesgerechtigde bevolking tussen 18 en 75 jaar. Het zijn tweetraps-steekproeven met gelijke selectiekansen voor de secundaire eenheden (individuen) (ISPO/PIOP, 1998).[2] De Vlaamse steekproef bestaat uit 2.099 eenheden, de Waalse steekproef telt 1.258 eenheden (Brussel wordt hier buiten beschouwing gelaten). Het meetmodel werd eerst geëxploreerd in een toevallig deel van de Vlaamse steekproef en vervolgens getest in het tweede deel.

De schaal waarvan de equivalentie zal onderzocht worden, is een gebalanceerde schaal bestaande uit drie positief en drie negatief verwoorde items die gevoelens van bedreiging door de migranten uitdrukken. De Nederlandstalige en Franstalige verwoordingen worden weergegeven in tabel 1. De gebruikte antwoordschalen zijn alle van de vorm 'helemaal oneens' (score 5), 'oneens' (score 4), 'noch eens, noch oneens' (score 3), 'eens' (score 4), 'helemaal eens' (score 5).[3] De uitspraken hebben betrekking op het zich al of niet bedreigd voelen door de aanwezigheid van migranten (of gastarbeiders), om economische (tewerkstelling) of culturele redenen.

Tabel 1: Nederlandstalige en Franstalige verwoording van de zes uitspraken

Items	Nederlands	Frans
Item 1 (-)	Migranten zijn over het algemeen niet te vertrouwen.	En général, on ne peut pas se fier aux étrangers.
Item 2 (-)	Gastarbeiders zijn een gevaar voor de tewerkstelling van de Belgen.	Les travailleurs immigrés sont une menace pour l'emploi des Belges.
Item 3 (-)	Moslims zijn een bedreiging voor onze cultuur en gebruiken.	Les musulmans sont une menace poeur notre culture et nos usages.
Item 4 (+)	De migranten dragen bij tot de welvaart van ons land.	Les immigrés contribuent à la prospérité de notre pays.
Item 5 (+)	De aanwezigheid van verschillende culturen verrijkt onze samenleving.	La présence de différents cultures enrichit notre société.
Item 6 (+)	Wij zouden de buitenlanders die zich in België vestigen hartelijk welkom moeten heten.	Nous devrions souhaiter de tout coeur la bienvenue aux étrangers qui veulent s'établir en Belgique.

4. Construct-equivalentie

In een recente studie hebben Billiet en McClendon (1999) aangetoond dat een model waarin, naast twee inhoudelijke factoren, een factor werd opgenomen die de tendens tot instemmen

meet, zowel in Vlaanderen als in Wallonië beter past bij de data dan een model waarin die antwoordstijl niet is opgenomen. De als derde factor gespecificeerde antwoordstijl correleert vrijwel perfect (méér dan 0,90) met een score voor instemming (aantal keren dat werd ingestemd met 14 gebalanceerde items). Wij beschouwen deze antwoordstijl als een methode-effect vermits deze het gevolg is van het gebruik van items van het Likert-type (eens-oneens). In de vroegere studie werd echter niet getest of de concepten equivalent zijn in de twee culturele contexten (en twee talen). In voorliggende studie wordt gebruik gemaakt van één gebalanceerde set, maar we zullen nu testen of het gemeten construct en de antwoordstijl equivalent zijn in de twee culturen.

De modellen worden getest met behulp van LISREL® 8.30 (Jöreskog & Sörbom, 1993). Aangezien het om ordinale gegevens gaat, wordt gebruik gemaakt van polychorische correlaties en van een gewogen kleinste-kwadratenschatting op basis van de asymptotische variantie-covariantiematrix. Voor het vergelijken van groepen kan het gebruik van polychorische correlaties tot verkeerde conclusies leiden indien de *thresholds* voor de variabelen in de twee groepen niet gelijk zijn. In de multigroup-vergelijking wordt er daarom voor gezorgd dat bij de berekening van de matrices de *thresholds* bij de Franstaligen gelijk zijn aan deze van de Vlamingen. Dit gebeurt met behulp van PRELIS™ 2 (Jöreskog & Sörbom, 1990).

In wat volgt wordt de identiteit van het meetmodel tussen de twee groepen onderzocht om op die wijze tot een conclusie te komen over de crossculturele equivalentie, rekening houdend met een methode-effect. Daarvoor wordt een aantal eisen gesteld aan het model.

De factorladingen van de items op het inhoudelijke construct (F_1) worden gelijk gesteld over de twee groepen (Vlamingen en Walen):

$$\lambda_{ij}^{1} = \lambda_{ij}^{2} \text{ (voor i} \in \{1,\ldots 6\} \text{ en j = 1)} \qquad (1)$$

Construct-equivalentie vereist immers dat het construct dat men wil onderzoeken in beide groepen op een voldoende gelijkaardige manier werd geoperationaliseerd. Rensvold en Cheung (1998) definiëren deze vraag naar construct-equivalentie operationeel als de vraag naar factoriële invariantie. Restrictie (1) impliceert de toetsing van de vraag naar factoriële invariantie en dus tevens van de vraag naar construct-equivalentie. Indien de factoriële invariantie van het construct over de onderzochte groepen kan worden aangetoond, mag met andere woorden besloten worden tot construct-equivalentie.

We moeten echter niet noodzakelijk aannemen dat de spreiding van de houding tegenover migranten (zich bedreigd voelen) gelijk is in de twee groepen. Dat is immers geen vereiste voor de equivalentie van de constructen. Toch zullen we starten met een model waarin ook deze twee varianties aan elkaar gelijkgesteld worden:

$$Var(F_1^1) = Var(F_1^2) \qquad (2)$$

We veronderstellen dat de zes indicatoren alle even sterk beïnvloed worden door de methode (acquiescence), en dus stellen we de ladingen op de stijlfactor (F_2) binnen de groepen aan elkaar gelijk:

$$\lambda_{1j}^{g} = \lambda_{ij}^{g} \text{ (voor i} \in \{2,\ldots 6\} \text{ en j = 2})} \quad (3)$$

We laten evenmin toe dat die ladingen verschillen tussen de twee groepen. Dit impliceert de veronderstelling dat de tendens tot instemmen in beide taalgroepen even sterk is (factoriële invariantie met betrekking tot de stijlfactor).

Tabel 2: Vergelijking van de meetmodellen met inhoudelijke factor en stijlfactor bij Vlamingen en Walen (steekproef van Vlamingen: N = 1.056; Walen = 1.068) (polychorischecorrelaties en WLS schattingen)[4]

Model	χ^2	*df*	*p*	RMSEA	*p*-value of close fit	$Var(F_1^g)$[a] dreiging	$Var(F_2^g)$[a] stijl
Model 1a	110,599	23	0,000	0,060	0,849	0,574	0,046
						(34,071)	(5,246)
alle parameters gelijk							
Model 1b.	89,972	22	0,000	0,054	0,949	0,634	
						(30,433)	0,028
$var(F_1^1) \neq var(F_1^2)$						0,548	(5,125)
						(27,672)	
Model 1c	83,044	21	0,000	0,053	0,953	0,619	0,048
						(29,365)	(6,044)
$var(F_1^1) \neq var(F_1^2)$;						0,548	0,025
$var(F_2^1) \neq var(F_2^2)$						(27,115)	(3,201)
Model 1d	63,008	20	0,000	0,045	0,991	0,611	0,048
						(29,113)	(6,106)
$var(F_1^1) \neq var(F_1^2)$;						0,537	0,019
$var(F_2^1) \neq var(F_2^2)$;						(26,784)	
error $cov(\varepsilon_1,\varepsilon_3)^2 \neq 0$							
Model 1[e]	44,523	19	0,001	0,036	0,999	0,621	0,048
$var(F_1^1) \neq var(F_1^2)$;						(29,133)	(6,063)
$var(F_2^1) \neq var(F_2^2)$;						0,527	0,018
error $cov(\varepsilon_1,\varepsilon_3)? \neq 0$;						(26,195)	(2,381)
$\lambda_{3,1}^{1} \neq \lambda_{3,1}^{2}$							
Model 2f	34,14	18	0,012	0,029	1,000	0,639	0,048
$var(F_1^1) \neq var(F_1^2)$;						(28,783)	(5,772)
$var(F_2^1) \neq var(F_2^2)$;						0,513	0,018
error $cov(\varepsilon_1,\varepsilon_3)? \neq 0$;						(24,995)	(2,229)
$\lambda_{3,1}^{1} \neq \lambda_{3,1}^{2}$;							
$\lambda_{2,1}^{1} \neq \lambda_{2,1}^{2}$							

[a] Kritische ratio's tussen haakjes. In geval van verschil heeft de bovenste parameter betrekking op Vlaanderen en de onderste op Wallonië.

$$\lambda_{ij}^{1} = \lambda_{ij}^{2} \text{(voor i} \in \{1,\ldots 6\} \text{ en j = 2)} \quad (4)$$

Dit laatste zal echter getest worden. Het is immers mogelijk dat er wel verschillen zijn in de tendens tot instemmen naargelang de culturele context. Ten slotte, veronderstellen we dat de varianties van de stijlfactor gelijk zijn in beide groepen.

$$Var(F_2^1) = Var(F_2^2) \quad (5)$$

Deze assumptie zal echter eveneens getest worden, aangezien het niet uitgesloten is dat de spreiding van het instemmen verschilt naargelang de culturele context.

Het model dat volledig beantwoordt aan de hierboven beschreven specificaties (1a) past niet bij de gegevens. Uit de verschillende tests blijkt duidelijk dat de varianties van de inhoudelijk factor en van de stijlfactor verschillen tussen de groepen (Model 1b en 1c). Bovendien moet in de Waalse steekproef een error-covariantie toegestaan worden tussen de eerste en de derde indicator (Model 1d). Dit volstaat nog niet. Ook de ladingen van de derde (Model 1e) en de tweede indicator (Model 1f) op het inhoudelijke latente construct zijn verschillend in de twee groepen. Model 1f is een passend model dat niet verder hoeft aangepast te worden.

Tabel 3: Parameters onder het passend model 2f

	Gestandaardiseerde (common metric) factorladingen		
	Inhoudelijke factor (Dreiging)		Methode factor (Instemming)
Items	Vlaanderen	Wallonië	Identiek
Item 1: Wantrouwen	0,795	0,795	0,178
Item 2: Tewerkstelling	*0,696*	*0,780*	0,178
Item 3: Cultuur	*0,604*	*0,730*	0,178
Item 4: Welvaart	-0,676	-0,676	0,178
Item 5: Cultureel rijk	-0,696	-0,696	0,178
Item 6: Welkom	-0,659	-0,659	0,178

We mogen besluiten dat de zes indicatoren globaal genomen hetzelfde construct meten in de twee groepen. De construct-equivalentie is met andere woorden gerealiseerd. Wel is het zo dat de tweede en derde indicator, beide met een negatieve verwoording, iets sterker door het construct bepaald worden in Wallonië dan in Vlaanderen. Het verschil tussen de factorladingen van beide taalgroepen is echter niet van dien aard dat men dient te besluiten tot construct-inequivalentie. Zowel voor het tweede als het derde item zijn de factorladingen in de twee taalgroepen immers van dezelfde ordegrootte. De hogere factorlading van het tweede item (bedreiging van tewerkstelling, zie tabel 1) bij de Waalse respondenten is echter niet verwonderlijk. Het werkloosheidsprobleem doet zich immers sterker gevoelen in het Waalse landsgedeelte dan in het Vlaamse. Gevoelens van bedreiging door migranten uiten zich daar-

om in Wallonië sterker op het vlak van de werkgelegenheid dan dit in Vlaanderen het geval is. Een verklaring voor de hogere factorlading van het derde item (bedreiging voor cultuur, zie tabel 1) bij de Waalse respondenten is echter minder evident. Misschien is het cultuurchauvinisme sterker aanwezig bij de Waalse bevolking dan bij de Vlamingen.

De tendens tot instemmen blijkt even sterk in Vlaanderen als in Wallonië. Er is met andere woorden wel sprake van een intracultureel methode-effect (de stijlfactor verklaart immers een significant deel van de gemeenschappelijke variantie, zie tabel 2), maar niet van een methode-effect in de crossculturele betekenis. Het effect van de methode is immers even sterk in Vlaanderen als in Wallonië.

5. Item-equivalentie

Item bias, als bron van item-inequivalentie, 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. Er bestaan verschillende technieken om een dergelijke vorm van bias te onderzoeken. Voor een bespreking van deze technieken verwijzen we naar Van de Vijver en Leung (1997). Om na te gaan in welke mate de items van de attitudeschalen in de beide taalgroepen onderhevig zijn aan *item bias*, maken we hier gebruik van covariantie analyses.

Daarbij wordt 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 waartoe de respondenten behoren. Het scoreniveau waartoe de respondent behoort, fungeert als covariaat in de analyse. De rationale achter deze werkwijze is immers dat personen met een zelfde score op het achterliggende concept (een zelfde scoreniveau), een zelfde score dienen te hebben op elk item van de etnocentrismeschaal ongeacht de culturele groep waartoe ze behoren. Het globale scoreniveau werd geconstrueerd op basis van de som van de scores van de respondenten op alle items van de etnocentrismeschaal, 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 vijfpuntenschaal. Indien de scores voor deze items worden opgeteld bekomt men de waarden 5 en 25 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 tot stand komen. De score 15 kan men bijvoorbeeld bekomen indien de respondent telkens de waarde 3 als antwoord geeft, maar ook door bijvoorbeeld tweemaal 5, tweemaal 2 en eenmaal 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 covariantie analyses fungeert een item van de etnocentrismeschaal 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 vijf items van de etnocentrismeschaal ook hoog scoren op het desbetreffende item. De twee andere effecten zijn relevanter voor onze analyse. 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 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 de taalgroep waartoe de respondent behoort. In deze analyse dient geen rekening gehouden te worden met de tendens tot instemmen, vermits uit de vorige paragraaf is gebleken dat beide culturele groepen in dezelfde mate vatbaar zijn voor dit methode-effect. Eventuele verschillen tussen Vlamingen en Walen op het niveau van de individuele items kunnen dus niet aan dit methode-effect worden toegeschreven.

Gezien de omvang van de steekproef lijkt het zinvol om een strikter significantieniveau te hanteren dan het gebruikelijke α-niveau van 5 procent. Bij een grote steekproefomvang zullen immers zelfs triviale, niet betekenisvolle verschillen tussen de twee groepen leiden tot significante waarden voor de teststatistiek. Vandaar dat voor de evaluatie van de equivalentie van de individuele items een significantieniveau van 0,01 wordt gehanteerd.

Tabel 4: Covariantie analyses voor de individuele items van de etnocentrismeschaal

		Hoofdeffect taalgroep		Interactie taalgroep/scoreniveau	
Items	N	*F*-waarde	R^2 [a] (%)	*F*-waarde	R^2 [a] (%)
Item 1	2.791	3,37	0,2	0,68	0,0
Item 2	2.802	6,63**	0,4	0,37	0,0
Item 3	2.795	6,37**	0,5	3,51	0,3
Item 4	2.782	0,30	0,0	0,48	0,0
Item 5	2.763	3,22	0,2	0,06	0,0
Item 6	2.790	2,45	0,2	1,04	0,1

* = p<0,05; ** = p<0,01

[a] kwadratensom van het effect gedeeld door de kwadratensom van het model (aandeel van het effect in de verklaarde variantie).

De resultaten van de item-bias-analyse worden weergegeven in tabel 4. Uit deze tabel blijkt dat geen enkel van de items van de etnocentrismeschaal beïnvloed wordt door een niet-uniforme vertekening. Immers, in geen enkel geval is de interactie tussen de taalgroep waartoe de respondent behoort en het globale scoreniveau statistisch significant. Dit laat een eenduidige interpretatie van de hoofdeffecten van de taalgroep toe.[5]

Voor twee van de zes items is het hoofdeffect van de taalgroep statistisch significant. Dit betekent dat deze twee items worden beïnvloed door een uniforme vertekening. Het aandeel

van deze vertekening in de verklaarde variantie van het model is echter onbeduidend laag. Voor beide items kan nauwelijks een half procent van de verklaarde variantie worden toegeschreven aan de uniforme vertekening. Het lijkt bijgevolg plausibel om te stellen dat de uniforme vertekening die beide items beïnvloedt, slechts weinig betekenisvol is en dat een rechtstreekse vergelijking van de scores van Vlamingen en Walen op de etnocentrismeschaal mogelijk is.

6. Besluit

In deze bijdrage werd een tussenliggend standpunt ingenomen op het continuüm tussen de idiografische en de nomothetische visie op cultuur en crosscultureel onderzoek, namelijk: crossculturele vergelijkingen zijn mogelijk op voorwaarde dat de validiteit van de vergelijking wordt aangetoond. De manier waarop men de validiteit van een crossculturele vergelijking kan aantonen, werd geïllustreerd aan de hand van een meetschaal voor etnocentrisme die aan een representatieve steekproef van Vlamingen en Walen werd voorgelegd. Concreet werd onderzocht of de constructen die in beide groepen gebruikt werden om het concept 'etnocentrisme' te meten in voldoende mate equivalent zijn. Hiervoor werd gebruik gemaakt van structural equation models, vermits het via deze techniek eveneens mogelijk was om te controleren voor een specifieke antwoordtendens, namelijk de tendens tot instemmen. Uit de toetsingen bleek dat niet alle factorladingen van de etnocentrismeschaal aan elkaar gelijk gesteld konden worden over de beide groepen heen. Globaal genomen waren er echter weinig substantiële verschillen tussen de factorladingen in de beide groepen en bijgevolg kan geconcludeerd worden dat de constructen in voldoende mate equivalent zijn (hoewel volledige invariantie verworpen dient te worden). Hieruit kan men concluderen dat men met behulp van de etnocentrismeschaal wel degelijk een zelfde construct meet bij Vlamingen en Walen. Met betrekking tot het methode-effect kan gesteld worden dat er wel degelijk sprake is van een methode-effect binnen elke groep. Anderzijds kan men echter niet spreken van een methode-effect in de crossculturele zin. Vlamingen en Walen bleken immers in dezelfde mate vatbaar voor de tendens tot instemmen.

De covariantie analyses voor de evaluatie van de item-equivalentie tonen aan dat ook op dit vlak kan besloten worden tot crossculturele equivalentie. Bij sommige items werd wel een significante item bias vastgesteld, maar bij nader toezicht bleek deze vertekening verwaarloosbaar. Er werden met andere woorden geen substantiële verschillen tussen de taalgroepen vastgesteld die te wijten zouden kunnen zijn aan enige anomalieën op het niveau van de individuele items. Op basis van bovenstaande bevindingen kan men besluiten dat het gebruikte meetinstrument geschikt is voor het maken van crossculturele vergelijkingen. Het is zelfs mogelijk om de scores van Vlamingen en Walen op de etnocentrismeschaal rechtstreeks met elkaar te vergelijken. Vermits de Vlamingen significant hoger scoren op de gesommeerde schaal dan de Waalse respondenten ($p<0{,}0001$), kan men dus besluiten dat de Vlamingen zich meer bedreigd voelen door migranten dan de Waalse respondenten.

Noten

1. Jerry Welkenhuysen-Gybels en Jaak Billiet zijn respectievelijk als aspirant van het Fonds voor Wetenschappelijk Onderzoek-Vlaanderen en gewoon hoogleraar verbonden aan de afdeling dataverzameling en -analyse van het departement sociologie van de Katholieke Universiteit Leuven.
2. Het Interuniversitair Steunpunt Politieke Opinieonderzoek is gevestigd aan de K.U.Leuven (ISPO)en de U.C.L. (PIOP) en wordt geleid door J. Billiet (promotor ISPO), M. Swyngedouw (onderzoeksdirecteur ISPO), A. Frognier (promotor PIOP) en A.M. Aish (onderzoeksdirecteur PIOP). De onderzoekers bij de verkiezingen van 1995 waren
3. A. Carton (interviewnetwerk ISPO), R. Beerten (data management ISPO) en Serge Van Diest (interviewernetwerk en datamanagement PIOP).
4. Voor deze analyse werden de schaalwaarden omgekeerd ten opzichte van de vragenlijst (1 = 5, enz...).
5. De aantallen in de steekproeven zijn lager omdat bij de gebruikte schattingsmethode er een uitsluiting is van alle observaties met ontbrekende waarden.
6. Eén van de assumpties in een covariantie analyse is immers 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, Auquier, Hauck, Oakes, Vandaele & Weisberg, 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). In dergelijke gevallen dient het hoofdeffect van de culturele groep op de score op een bepaald item nader bekeken te worden (Welkenhuysen-Gybels, 1998).

Literatuur

Anderson, S., Auquier, A., Hauck, W.W., Oakes, D., Vandaele, W., Weisberg, H. (1980). *Statistical methods for comparative studies. Techniques for bias reduction.* New York: John Wiley & Sons.

Berry, J.W., Poortinga, Y.H., Segall, M.H. & Dasen, P.R. (1992 (1993)). *Cross-cultural psychology. Research and applications.* Cambridge: Cambridge University Press.

Billiet, J.B. & McClendon, M.J. (1998). On the identification of acquiescence in balanced sets of items using structural models. In: Ferligoj, A. (Ed.) *Advances in methodology, data analysis and statistics* (pp. 129-150). Ljubljana: FDV.

Billiet, J.B. & McClendon, M.J. (2000). Modelling aqcuiescence in measurement models for two balanced sets of items. *Structural Equation Modelling* (aanvaard voor publicatie).

ISPO/PIOP (1998). *General Election Study Belgium. Codebook and Questionnaire.* Leuven : ISPO/PIOP, K.U. Leuven/UCL.

Johnson, T.P. (1998). Approaches to equivalence in cross-cultural and cross-national survey-research. *ZUMA Nachrichten Spezial, Cross-cultural survey equivalence, No. 3,* 1-40.

Jöreskog, K.G. (1990). New developments in LISREL. Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and quantity, 24,* 387-404.

Jöreskog, K.G. & Sörbom, D. (1993). *LISREL® 8 Users's Reference Guide.* Chicago: Scientific Software International.

Rensvold, R.B. & Cheung, G.W. (1998). Testing measurement models for factorial invariance: a systematic approach. *Educational and psychological measurement, 58,* 1017-1034.

Vijver, F.J.R. van de (1998). Towards a theory of bias and equivalence. *ZUMA Nachrichten Spezial, Cross-cultural survey equivalence, No. 3,* 41-65.

Vijver, F.J.R. van de & Hutschemaekers, G.J.M. (1990). *The investigation of culture. Current issues in cultural psychology.* Tilburg: Tilburg University Press.

Vijver, F. van de & Leung, K. (1997), *Methods and Data Analysis for Cross-Cultural Research.* Thousand Oaks: Sage.

Welkenhuysen-Gybels, J. (1997). Van de Vijver, Fons & Leung, Kwok, Methods and Data Analysis for Cross-Cultural Research. Thousand Oaks, Sage, 1997, 200 p., boekbespreking, in: *Tijdschrift voor Sociologie, 18,* 523-525.

Welkenhuysen-Gybels, J. (1998). Cross-cultureel onderzoek: methodologische problemen bij de toepassing van een verkorte F-schaal in Franstalig en Nederlandstalig België. *Tijdschrift voor Sociologie, 19,* 449-472.