



**Nationaal Onderzoeksinstituut  
voor Arbeidsomstandigheden**

**Validering van de Kern-VBBA in Vlaanderen**

**Notelaers, Guy** (Nationaal Onderzoeksinstituut  
Voor Arbeidsomstandigheden, K.U. Leuven)

**Dr. van Veldhoven, Marc** (Stichting  
Kwaliteitsbevordering Bedrijfsgezondheidszorg,  
K.U. Nijmegen)

## **Voorstelling van de data uit Vlaanderen**

De eerste reeks van zeven bedrijven leverde 1013 vragenlijsten op van de kern-VBBA. Deze observaties komen uit verschillende sectoren:

- Interimsector
- Farmaceutische sector
- Bouwsector
- Gezondheidszorg
- Petrochemie
- Chemische sector afgeleide producten
- Petrochemie afgeleide producten

Als we het hebben over voldoende heterogeniteit kunnen we op verschillende manieren de dataset evalueren. Naar sector gezien hebben we een allegaartje : uitzendsector, farmacie, petrochemie, bouw, verpleging. Als we kijken naar de aard van het werk zien we echter dat we een oververtegenwoordiging van bedienden (+/- 600) hebben ten opzichte van een ondervertegenwoordiging van arbeiders (+/- 330).<sup>1</sup> De kaders of hogere bedienden zijn met ongeveer 75. Het geslacht is zowel in de lage als in de hoge functies gelijk verdeeld.

## **Schaalbaarheid en betrouwbaarheid vergeleken met Nederlandse ontwikkelingsgroep**

Om de gegevens met Mokkens Scaling Program te bewerken moeten we helaas heel wat data deleten. Na listwise deletion houden we 642 observaties over.

In dat programma gebruiken we niet de search procedure die, zoals de Engelse term het zegt, gericht is op de schaalconstructie met bepaalde voorwaarden ((h)((i)j)=cte=.3). We gebruiken 'type=test', een procedure die erop gericht is om de items en itemparen te testen op hun schaalbaarheidskwaliteit. Daarbij gaat het programma ervan uit dat de aangereikte items schaalbaar zijn, waardoor we onze aandacht kunnen richten op de schaalbaarheidscoëfficiënt

---

<sup>1</sup> Er is ook een fikse afwijking met de samenstelling van de ontwikkelingsgroep in het doctoraat van van Veldhoven. Deze verklaart wellicht ook de lage Hi waarde van het derde item bij de schaal 'lichamelijke inspanning'.

van de hele schaal nl. Loevinger's H. Deze moet in eerste instantie groter zijn dan .3 volgens de literatuur.<sup>2</sup>

Hieronder geven we niet alleen een overzicht over deze kwaliteitsmaat maar geven we ook het verschil met de Nederlandse ontwikkelingsgroep aan. Daarnaast berekenden we ook de betrouwbaarheid van de schaal wanneer we ze opnieuw zouden inzetten aan de hand van Chronbach's alfa en het verschil van deze maat met de betrouwbaarheid in de Nederlandse ontwikkelingsgroep. Tenslotte bekijken we in de laatste kolom de schendingen van de dubbele monotonie. Over de monotone homogeniteit of de eis dat we één latente trek hebben, rapporteren we bijgevolg niet.<sup>3</sup>

Schalen	Loevinger 's weighted H	Afwijking t.o.v. Nederland	Cronbach's alfa	Afwijking ten opzichte van Nederland	DM Aantal schendingen in P matrix boven 10% per item geïnspecteerd
Werktempo en hoeveelheid	.47	-.01	.86	-.02	0
Emotionele belasting	.43	-.02	<b>.77</b>	-.06	0
Lichamelijke inspanning	.44 <sup>4</sup>	-.12	.87	-.03	Item 3
Afwisseling in het werk	.44	.02	<b>.75</b>	-.04	0
Leermogelijkheden	.72	.15	.85	.01	0
Zelfstandigheid in het werk	.59	.15	.92	.02	0
Relatie collega's	.56	.04	.87	.01	0
Relatie directe leiding	.6 <sup>5</sup>	.02	.89	.01	0
Inspraak	.61	.19	.89	.04	0
Toekomstonzekerheid	.84	.09	.95	.01	0
Plezier in het werk	.52	.03	.82	.05	0
Betrokkenheid bij de organisatie/ het bedrijf	.47	-.07	<b>.75</b>	-.01	0
Herstelbehoefte	.54	.06	.88	.01	0
Piekeren	.64	-.03	<b>.75</b>	.01	0
		+.05/14 = .04		.01/14=.00	

Uit de tabel blijkt dat het met de schaalbaarheid van de kern-vbba-items erg goed zit. De laagste schaalbaarheid vinden we bij de items die over emotionele belasting gaan. De hoogste

<sup>2</sup> Mokken, R.J. Nonparametric Models for Dichotomous Responses. In : van der Linden, W.J. Hambleton, R.K. Handbook of Modern Item Response Theory. 1997, p. 361.

<sup>3</sup> Het testen van de dubbele monotonie impliceert het hebben van een monotoon mokkenmodel.

<sup>4</sup> Item 3 schendt de dubbele monotonie

<sup>5</sup> item 4 schendt monotonie maar  $H_i = .38$

schaalbaarheid vinden we bij toekomstonzekerheid. Dat betekent dat we volgens Mokken in België te maken hebben met mediumschalen ( $.4 < H < .5$ ) zoals werktempo en hoeveelheid, emotionele belasting, lichamelijke inspanning en afwisseling in het werk. De overige schalen in de kernvragenlijst zijn sterke schalen ( $H > .5$ ). Voor het 'makkelijke' Mokkenmodel moeten we vermelden dat bij relatie met directe leiding een item in de search-procedure uitgesloten zou worden nl. 'Heeft u conflicten met uw directe leiding?' Technisch gezien is dat te wijten aan de schaalbaarheid van de itemparen Hij die in sommige gevallen beneden de threshold van .3 is. Concreet betekent het dus dat dit item niet monotoon stijgend is.

Wanneer we H vergelijken met de H uit van Veldhoven zijn ontwikkelingsgroep, dan kunnen we globaal stellen dat de Belgische data en dito schalen gemiddeld een H hebben die .04 hoger ligt. Schaalgewijs zijn er echter wel enkele grote afwijkingen waarvan de meesten echter positief zijn. Aberrant daarin is de schaal 'lichamelijke inspanning' (cf. infra). In de ontwikkelingsgroep komt deze schaal naar voor als een sterke schaal terwijl in de Belgische sample ze als mediumschaal op het voorplan treedt. Dezelfde verschuiving zien we echter ook bij 'betrokkenheid bij de organisatie' doch het absolute verschil is veel kleiner. In positieve zin wijken leermogelijkheden, zelfstandigheid in het werk en inspraak van de H-waarden in de Nederlandse ontwikkelingsgroep opvallend af. Hoewel we hier en daar wat afwijkingen vaststellen ten opzichte van de ontwikkelingsgroep in Nederland, zit het dus met de schaalkwaliteit van de Belgische kern-vbba goed.

Het strenge Mokkenmodel kunnen we op verschillende manieren inspecteren. Volgens Sijtsma, Debets en Molenaar kunnen we bij polytome items best 'check=restscore' gebruiken terwijl voor dichotome items 'check=htrans' de beste methode is om te kijken of de item-responscurven mekaar (significant) snijden. De default optie in MSP stelt dat van een schending sprake is als in 2% van de gevallen de IRFs mekaar snijden. Dat is gezien het onderzoeksveld (we houden ons niet bezig met toetsontwikkeling) waarin wij opereren, nl. dat van de werkstress, erg streng. Daarom stellen we voor om het bij een schendingsniveau te houden met 10% zoals in het doctoraat van Veldhoven. Om de vergelijkbaarheid echter te bevorderen met de ontwikkelingsgroep opteren we er voor om zowel voor polytome en dichotome items 'check=p-matrix' te nemen. Dat betekent : de dubbele monotonie nagaan aan de hand van het percentage schendingen in de P-Matrices. In de laatste kolom van de tabel (cf. supra) rapporteren we maar één schending boven 10%. Het gaat om het item 'Heeft u tijdens het werk veel hinder van regelmatig te hoog te moeten reiken' in de schaal

lichamelijke inspanning. Er is echter veel te vinden om dat item te laten in de Vlaamse versie.<sup>6</sup> We kunnen immers verwijzen naar het feit dat arbeiders ondervertegenwoordigd zijn in de dataset uit Vlaanderen. Maar we verliezen niets als we het item herformuleren. In de versies van de VBBA die vanaf december door Quest verspreid worden, zullen we een geherformuleerd item geven aan de respondenten.

Omdat het maar om één schaal gaat, die dan nog niet in de VBBA-structuur opgenomen is door de confirmatorische factoranalyse, kunnen we zeggen dat de Kern-VBBA in Vlaanderen voldoet aan de eis van de dubbele monotonie. De schalen in Vlaanderen zijn dus unidimensioneel en de item responscurven zijn niet snijdend monotoon stijgend.

Wanneer we de betrouwbaarheid van de schalen nagaan kunnen we zeggen dat de individuele schaalverschillen tussen de Belgische data en de Nederlandse ontwikkelingsdata erg klein zijn. Globaal gezien zitten we op dat vlak goed. Toch zien we dat vier schalen een betrouwbaarheid van minder dan 80% hebben. De enige schaal die 'normatief' verschilt van de Nederlandse ontwikkelingsgroep is emotionele belasting (norm : chronbachs alfa minstens .8). Volgens de gangbare literatuur zouden we geneigd zijn om te zeggen dat dat te laag is. In dezelfde literatuur zien we echter wel dat we omzichtig moeten omgaan met deze 'afgesproken normen'. 'Wanneer de vaardigheid met meer toetsen of op verschillende momenten wordt getoetst, zijn lagere ondergrenzen acceptabel, waarbij in de literatuur .65 wel als gewenste ondergrens wordt genoemd'<sup>7</sup>.

## **Besluit**

De vergelijking van de 642 Vlaamse observaties uit de Kern-VBBA met de Nederlandse ontwikkelingsgroep op het gebied van schaalkwaliteit en betrouwbaarheid levert erg goede resultaten op. We zien geen enkele reden om de Kern-vragenlijst te wijzigen. De schalen voldoen aan de eis van de dubbele monotonie en de betrouwbaarheid van het instrument is erg groot. De Kern-Vragenlijst kan in Vlaanderen worden toegepast met vergelijkbare schaalkwantiteiten als in Nederland. Wel is de opbouw van een eigen referentiebestand voor Vlaanderen van belang, i.v.m. de mogelijke culturele specificiteit van de problematiek van psycho-sociale arbeidsbelasting en werkstress.

---

<sup>6</sup> zie : voetnootnummer 1

<sup>7</sup> Veldhuijzen, N.H. Goldebeld, P. Sanders P.F. Klassieke testtheorie en generaliseerbaarheidstheorie. In : Eggen, T. Sanders, P.F. Psychometrie in de praktijk. Cito Instituut voor Toetsontwikkeling. Arnhem, 1993, p. 61.