

Psychometrische eigenschappen van meetinstrumenten in dagboekonderzoek

Illustratie aan de hand van bevrologenheid*

Kimberley Breevaart, Arnold B. Bakker & Evangelia Demerouti**

Ondanks de opkomende populariteit van dagboekonderzoek wordt er tot op heden weinig aandacht besteed aan de psychometrische eigenschappen van de meetinstrumenten die in dergelijk onderzoek worden gebruikt. Wanneer in dagboekonderzoek het tijdsframe van vragenlijsten wordt aangepast, is het belangrijk om te onderzoeken of de factorstructuur hetzelfde blijft. Het doel van dit onderzoek is het introduceren van een methode om de factorstructuur van dagelijkse variabelen te onderzoeken. Dit zal worden toegelicht door een multiniveau confirmatieve factoranalyse uit te voeren op de vragenlijst om dagelijkse bevrologenheid te meten. Uit de resultaten blijkt dat dagelijkse bevrologenheid gemeten kan worden door het tijdsframe aan te passen van de vragenlijst om langdurige bevrologenheid te meten, waarbij het wenselijk is om één van de negen items te vervangen.

1 Introductie

Bolger, Davis en Rafaeli schreven in 2003 een theoretisch artikel met als titel 'Diary methods, capturing life as it is lived'. Hiermee willen de auteurs duidelijk maken dat dagboekmethoden ervaringen en gebeurtenissen kunnen onderzoeken in hun natuurlijke context, in tegenstelling tot en in aanvulling op meer traditionele meetmethoden. Door één of meer keren per dag korte vragenlijsten af te nemen kunnen ervaringen en gebeurtenissen onderzocht worden die sterk afhankelijk zijn van de situatie of die fluctueren over en/of binnen dagen. Vergeleken met traditionele methoden zoals vragenlijstonderzoek met een cross-sectioneel of longitudinaal onderzoekontwerp, waarin respondenten wordt gevraagd om terug te kijken over enkele weken/maanden, wordt de retrospectieve vertekening in dagboekonderzoek drastisch gereduceerd.

* Dit artikel is met toestemming van de uitgever deels gebaseerd op Breevaart, Bakker, Demerouti & Hetland (2012).

** Kimberley Breevaart is verbonden aan de Erasmus Universiteit Rotterdam. Correspondentieadres: Erasmus Universiteit Rotterdam, Vakgroep Arbeids- en Organisationspsychologie, Postbus 1738, 3000 DR Rotterdam, tel. 010-4089034, e-mail breevaart@fsw.eur.nl. Arnold B. Bakker is verbonden aan de Erasmus Universiteit Rotterdam. Evangelia Demerouti is verbonden aan de Technische Universiteit Eindhoven.

Reviews van Wheeler en Reis (1991) en Reis en Gable (2000) laten zien dat dagboekonderzoek steeds meer wordt gebruikt in verschillende domeinen. Het is daarom des te opvallender dat er tot nu toe weinig aandacht is geschonken aan de psychometrische eigenschappen van constructen die in dagboekonderzoek worden gemeten. Dit is echter wel belangrijk, omdat het tijdsframe van items om algemene variabelen te meten (bijv. bevlogenheid) vaak wordt aangepast om dagelijkse variabelen te meten (Allen & Potkay, 1981). Recentelijk hebben Sonnentag, Dormann en Demerouti (2010) hierbij enkele kanttekeningen geplaatst. Deze kanttekeningen hebben betrekking op de factorstructuur van dagelijkse variabelen wanneer deze worden gemeten met aangepaste vragenlijsten die zijn ontwikkeld om algemene variabelen te meten.

In het huidige artikel zal een methode worden toegelicht om de psychometrische eigenschappen van constructen die worden gebruikt in dagboekonderzoek te onderzoeken. Dit wordt gedaan door een multiniveau confirmatieve factoranalyse (MCFA) uit te voeren op de Utrechtse Bevlogenheidsschaal (UBES; Schaufeli & Bakker, 2003, 2010) waarvan het tijdsframe is aangepast om dagelijkse bevlogenheid te meten. Voordat we hier verder op ingaan, zullen we eerst dieper ingaan op verschillende vormen van dagboekonderzoek met de bijbehorende voor- en nadelen, om duidelijk te maken welke verschillen er zijn met meer traditionele meetmethoden.

2 Dagboekonderzoek

2.1 Soorten dagboekonderzoek

Er zijn verschillende soorten dagboekonderzoek, die in drie brede categorieën ingedeeld kunnen worden: *signal-contingent*-methoden, *event-contingent*-methoden en *interval-contingent*-methoden (Wheeler & Reis, 1991). *Signal-contingent*-methoden, ook wel *experience-sampling*-methoden genoemd, zijn methoden waarbij respondenten gevraagd wordt een korte vragenlijst in te vullen wanneer ze daarvoor een signaal (bijv. e-mail/sms) krijgen. Dit signaal kan zowel op een vast of een willekeurig tijdstip worden aangeboden of op een combinatie van beide tijdstippen. Met *event-contingent*-methoden wordt aan respondenten gevraagd om een korte vragenlijst in te vullen elke keer dat een van tevoren bepaalde situatie zich voordoet, bijvoorbeeld wanneer ze een conflict hebben gehad met hun leidinggevende. Hiermee kunnen gebeurtenissen en ervaringen worden gemeten die zich zelden voordoen. *Interval-contingent*-methoden, ook wel dagelijkse dagboeken genoemd, worden gebruikt om op één of meer vaste tijdstippen per dag een vragenlijst af te nemen die niet gerelateerd is aan specifieke gebeurtenissen van die dag. Dit is ook de methode die gebruikt wordt in het huidige onderzoek.

Voordelen van deze dagboekmethoden zijn de vermindering van retrospectieve vertekening, de verzameling van data in de natuurlijke omgeving, en de mogelijkheid tot het onderzoeken van fluctuerende gevoelens, gedachten en gedragingen. Een nadeel ten opzichte van meer traditionele typen onderzoek is dat dit type onderzoek erg invasief kan zijn, vooral wanneer er meerdere dagen op meerdere tijdstippen per dag wordt gemeten. Het is dan ook aan te raden om de vragenlijst-

ten zo kort mogelijk te houden. Voor een gedetailleerd overzicht van de verschillende dagboekmethoden, de voor- en nadelen van deze methoden en het omgaan met de problemen die men tegenkomt in dit type onderzoek, wordt verwezen naar Bolger, Davis en Rafaeli (2003), en Bakker en Daniels (in press).

2.2 *Dagboekonderzoek naar bevlogenheid*

Een construct dat steeds vaker ook door middel van dagboekonderzoek wordt onderzocht, is bevlogenheid. Bevlogenheid is een positieve, werkgerelateerde toestand die wordt gekenmerkt door vitaliteit, toewijding en absorptie (Bakker, 2009; Schaufeli & Bakker, 2010). Vitaliteit wil zeggen dat iemand veel energie en mentale veerkracht heeft en bereid is om moeite te steken in het werk. Toewijding houdt in dat iemand erg betrokken is bij het werk en hier enthousiast over is. Absorptie betekent dat iemand volledig opgaat in het werk, erg geconcentreerd is op het werk en het gevoel heeft dat de tijd vliegt tijdens het werken. Dagelijkse bevlogenheid gaat over kortdurende ervaringen van bevlogenheid die fluctueren binnen personen over een korte periode (Sonnentag et al., 2010) en kan verklaren waarom bevlogen mensen ook wel eens een 'off-day' hebben. Dagboekonderzoek naar bevlogenheid laat inderdaad zien dat ongeveer 40% van de variantie in bevlogenheid verklaard kan worden binnen personen (bijv. Sonnentag, 2003; Xanthopoulou & Bakker, in press).

Zoals eerder is vermeld, wordt op het gebied van bevlogenheid steeds meer onderzoek gedaan door middel van dagelijkse dagboeken. Xanthopoulou, Bakker, Heuven, Demerouti en Schaufeli (2008) vonden in hun onderzoek onder stewardessen bijvoorbeeld dat dagelijkse sociale steun van collega's positief en direct was gerelateerd aan dagelijkse bevlogenheid tijdens de vlucht en indirect via het gevoel van eigen effectiviteit. In een ander onderzoek van Xanthopoulou, Bakker, Demerouti en Schaufeli (2009b) werd gevonden dat medewerkers in fastfoodrestaurants meer bevlogen waren op dagen dat ze veel werkgerelateerde en persoonlijke hulpbronnen hadden (o.a. dagelijkse autonomie, coaching, eigen effectiviteit en optimisme). Uit beide onderzoeken kwam ook naar voren dat medewerkers beter presteerden op dagen dat ze meer bevlogen waren. Uit onderzoek van Bakker en Bal (2010) bleek dat deze relaties ook over het tijdsbestek van een week standhielden. Leraren bleken meer bevlogen te zijn in weken waarin ze veel werkgerelateerde hulpbronnen (autonomie, contact met de leidinggevende, en ontplooiingsmogelijkheden) tot hun beschikking hadden. Zij bleken in die weken ook beter te presteren. Aangezien dagboekonderzoek steeds meer wordt gebruikt, is het belangrijk om de psychometrische eigenschappen te onderzoeken van de meetinstrumenten die daarbij worden gebruikt. Meten we in dagboekonderzoek daadwerkelijk wat we beogen te meten? Hier wordt op dit moment echter nog weinig aandacht aan besteed door onderzoekers. In het huidige onderzoek wordt een eerste stap in de juiste richting gezet door de factoriële validiteit van dagelijkse bevlogenheid te onderzoeken.

2.3 *Meetinstrumenten in dagboekonderzoek*

De vragenlijst voor het meten van dagelijkse bevlogenheid in dagboekonderzoek is vaak een aanpassing van de Utrechtse Bevlogenheid Schaal (UBES). De UBES is

oorspronkelijk ontwikkeld om langdurige (algemene) bevlogenheid te meten. Voor het meten van dagelijkse bevlogenheid wordt het tijdsframe van de originele schaal vaak aangepast (bijv. Sonnentag, 2003; Xanthopoulou et al., 2009a). Een voorbeeld hiervan is het item 'Op mijn werk bruis ik van energie', dat in dagboekonderzoek wordt gewijzigd in 'Vandaag bruiste ik van energie op mijn werk' (zie tabel 1 voor een overzicht van alle items).

Zoals eerder is aangegeven, hebben Sonnentag en haar collega's (2010) enkele kanttekeningen geplaatst bij de factorstructuur van de UBES wanneer deze aangepast wordt voor dagboekonderzoek. Een van deze kanttekeningen betreft de mate waarin dagelijkse bevlogenheid gelijk is aan algemene bevlogenheid. Wanneer dagelijkse en algemene bevlogenheid niet gelijk zijn, heeft dit gevolgen voor de factoriële invariantie en de factoriële validiteit van de aangepaste vragenlijst om dagelijkse bevlogenheid te meten. Het is bijvoorbeeld denkbaar dat de UBES naar gevoelens en/of attitudes vraagt die zich niet ontwikkelen of veranderen van dag tot dag, maar stabiel zijn of langzaam veranderen. Dit zou betekenen dat bevlogenheid niet op dezelfde manier gemeten kan worden op dagelijks en algemeen niveau en er wellicht een andere theoretische definitie nodig is van dagelijkse bevlogenheid.

Een andere kanttekening is dat sommige dimensies van bevlogenheid wellicht belangrijker zijn op bepaalde dagen. Het kan bijvoorbeeld zo zijn dat het werk alleen inspirerend is wanneer er tijdens een maandelijkse vergadering nieuwe ideeën worden gegenereerd. Dit kan resulteren in minder geëxtraheerde factoren in een factoranalyse wanneer het item over inspiratie niet op dagelijkse basis beantwoord kan worden. Gerelateerd hieraan is een ander algemeen probleem met het meten van dagelijkse variabelen, namelijk dat mensen consistent willen zijn in hun gedrag en geen gedrag willen rapporteren dat inconsistent is (Visser, Krosnick & Lavrakas, 2000). Doordat mensen consistente antwoorden geven, kunnen mogelijke echte veranderingen over de tijd onzichtbaar blijven. Dit lijkt echter onwaarschijnlijk, aangezien gemiddeld 42% van de variantie van bevlogenheid verklaard kan worden binnen personen (Xanthopoulou & Bakker, in press). Tot slot kan het zijn dat een van de vitaliteitsitems niet goed werkt op dagelijkse basis, omdat deze naar de ochtend verwijst ('Toen ik vanmorgen opstond, had ik zin om naar mijn werk te gaan'). Wanneer deze vraag aan het einde van de dag wordt gesteld, hoeft het geen reflectie te zijn van de hoeveelheid vitaliteit die wordt ervaren gedurende de dag.

In dit artikel wordt een multiniveau confirmatieve factoranalyse uitgevoerd op de aangepaste UBES om te onderzoeken of dagelijkse bevlogenheid met dezelfde factoren kan worden gemeten als algemene bevlogenheid. Dit is van belang voor de interpretatie van bestaand dagboekonderzoek naar bevlogenheid en voor dagboekonderzoek naar bevlogenheid in de toekomst. Dit is ook van belang voor dagboekonderzoek in het algemeen, omdat de psychometrische eigenschappen van dagelijkse variabelen tot nu toe weinig zijn onderzocht. Hoewel de meeste dagboekonderzoeken een goede interne consistentie van de subschalen van de UBES laten zien, is de factorstructuur van de dagelijkse versie van de UBES nog nooit onderzocht met multiniveau analyses. De 3-factorenstructuur van de 9-itemversie van de UBES is aangetoond in veel verschillende landen en beroepen, maar de

drie factoren (vitaliteit, toewijding en absorptie) blijken sterk aan elkaar gerelateerd (Schaufeli & Bakker, 2010; Schaufeli, Bakker & Salanova, 2006). Daarom zal in dit artikel zowel de 3-factorenstructuur als de 1-factorstructuur worden getoetst.

3 Methode

3.1 Procedure

Voor dit onderzoek is gebruikgemaakt van drie verschillende onderzoeken die dagelijkse bevoegenheid als een van de variabelen onderzochten. Deze data is verzameld door drie masterstudenten die studeren aan een Nederlandse universiteit. In deze onderzoeken werd de relatie tussen activiteiten in de vrije tijd en bevoegenheid onderzocht, de relatie tussen leiderschap en bevoegenheid, en de cross-over van bevoegenheid tussen collega's. Alle respondenten vulden aan het eind van de werkdag het dagboek in gedurende vijf opeenvolgende werkdagen. De gerapporteerde bevoegenheid aan het eind van elke werkdag reflecteert dagelijkse bevoegenheid, terwijl het gemiddelde over deze vijf dagen een reflectie is van algemene bevoegenheid.

3.2 Respondenten

De steekproef bestond uit 271 werknemers: 159 mannen en 112 vrouwen. De jongste werknemer was 20 en de oudste 64 ($M = 36.75$, $SD = 10.49$). De meeste werknemers waren getrouwd, (62.4%) en hoogopgeleid (72.3%). Ze werkten in de zakelijke dienstverlening (23.2%), bij de overheid (26.6 %) of in een andere sector, zoals onderwijs, gezondheidszorg en transport.

3.3 Meetinstrumenten

De Nederlandse versie van de 9-item UBES (Schaufeli et al., 2006) werd gebruikt om bevoegenheid te meten. Onderzoek heeft de validiteit en betrouwbaarheid van deze korte versie van de UBES aangetoond (Schaufeli et al., 2006). De vragen waren aangepast om dagelijkse bevoegenheid te meten (zie tabel 1). De vragen werden gemeten op een 7-puntsschaal (0 = geheel mee oneens, 6 = geheel mee eens). Voorbeelditems zijn: 'Vandaag inspireerde mijn werk mij' en 'Vandaag was ik erg enthousiast over mijn werk'. De interne consistentie van de totale schaal over alle dagen is .93, met een bereik van .92 tot .95. Voor absorptie is de interne consistentie over alle dagen .81 met een bereik van .76 tot .85. Vitaliteit heeft een interne consistentie van .87 over alle dagen met een bereik van .86 tot .89, en toewijding heeft een interne consistentie van .89 over alle dagen met een bereik van .86 tot .91.

3.4 Data-analyse

Wanneer onderzoekers naar de factorstructuur van een construct op persoonsniveau kijken, voeren ze meestal een factoranalyse uit die is gebaseerd op de totale covariantiematrix van de gehele steekproef, of een factoranalyse die is gebaseerd op de tussenpersoonsmatrix van de steekproef (Dyer, Hanges & Hall,

2005). Volgens Muthén (1994) zijn beide procedures echter problematisch. Om deze problemen tegen te gaan ontwikkelde Muthén de multiniveau confirmatieve factoranalyse, waarbij gelijktijdig gebruik wordt gemaakt van de tussen- en binnenpersoonsmatrix. Om deze analyse uit te voeren wordt gebruikgemaakt van Mplus (Muthén & Muthén, 2002).

De χ^2 , Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA) en Standardized Root Mean Square Residual (SRMR) zijn gebruikt om de fit van de modellen op de data te bepalen. CFI-waarden van .90 tot .95 zijn acceptabel, en waarden van .95 tot 1.00 zijn hoog. RMSEA-waarden lager dan .08 en .05 zijn respectievelijk acceptabel en goed, en SRMR-waarden tot .10 indiceren acceptabele fit (Schweizer, 2010). Aangezien de steekproef en het statistische onderscheidingsvermogen in dit onderzoek vrij groot is (271 deelnemers \times 5 dagen = 1.355 datapunten op het tussenpersoonsniveau) en de χ^2 wordt beïnvloed door de steekproefgrootte, ligt de focus in dit onderzoek op de andere fit-indices. Het Bayesian Information Criterion (BIC) werd gebruikt om de fit van de verschillende modellen te vergelijken. Aangezien de BIC corrigeert voor het aantal parameters dat geschat wordt (Schwartz, 1978), kan de BIC gebruikt worden om modellen te vergelijken die niet genest zijn. Een lagere BIC-waarde indiceert een betere fit.

4 Resultaten

4.1 Beschrijvende statistieken

Tabel 1 toont de gemiddelden, standaarddeviaties en intercorrelaties voor de negen items van de UBES op het tussen- en binnenpersoonsniveau. Zoals is af te lezen uit de tabel, zijn de correlaties op binnenpersoonsniveau wat lager dan op tussenpersoonsniveau. Dat wil zeggen dat de discriminante validiteit op binnenpersoonsniveau in mindere mate is toe te schrijven aan de gebruikte meetmethode dan op tussenpersoonsniveau.

4.2 Multiniveau confirmatieve factoranalyse

We hebben de intraklasse correlatiecoëfficiënt (ICC) gebruikt om te bepalen of het gebruik van multiniveau analyses gerechtvaardigd was. De ICC geeft aan hoeveel tussenpersoonvariantie er is in verhouding tot de totale variantie. In dit onderzoek varieerden de ICC's van .36 tot .55, waardoor ze als normaal geïnterpreteerd kunnen worden (James, 1982). Dit wil zeggen dat er voldoende variantie is op beide niveaus en het gebruik van multiniveau analyses gerechtvaardigd is.

Tabel 1 Gemiddelden, standaarddeviaties, tussen- en binnenpersoonsniveau-intercorrelaties voor de negen dagelijkse bevlogenheidsitems (N = 271)

Items	M	SD	1.	2.	3.	4.	5.	6.	7.	8.
1. Vandaag bruieste ik van energie op mijn werk. (Vi)	5.01	1.22								
2. Vandaag voelde ik me fit en sterk toen ik aan het werk was. (Vi)	5.09	1.23	.94 (.77)							
3. Toen ik vandaag opstond, had ik zin om aan het werk te gaan. (Vi)	4.93	1.42	.82 (.51)	.83 (.51)						
4. Vandaag was ik enthousiast over mijn baan. (Tw)	5.13	1.26	.84 (.57)	.88 (.59)	.86 (.41)					
5. Vandaag inspireerde mijn werk mij. (Tw)	4.88	1.32	.80 (.56)	.87 (.55)	.83 (.40)	.93 (.63)				
6. Vandaag was ik trots op het werk dat ik doe. (Tw)	5.09	1.29	.75 (.45)	.77 (.47)	.71 (.38)	.89 (.56)	.85 (.51)			
7. Toen ik vandaag heel intensief aan het werk was, voelde ik mij gelukkig. (Ab)	5.10	1.23	.79 (.49)	.82 (.52)	.81 (.46)	.92 (.55)	.87 (.54)	.87 (.59)		
8. Vandaag ging ik helemaal op in mijn werk. (Ab)	4.72	1.41	.62 (.53)	.66 (.51)	.64 (.39)	.81 (.55)	.74 (.54)	.82 (.61)	.76 (.53)	
9. Vandaag bracht mijn werk mij in vervoering. (Ab)	3.91	1.48	.59 (.36)	.56 (.36)	.60 (.24)	.65 (.40)	.69 (.38)	.70 (.42)	.66 (.39)	.69 (.53)

Vi = vitaliteit, Tw = toewijding, Ab = absorptie. Binnenpersoonsniveau-intercorrelaties staan tussen haakjes.

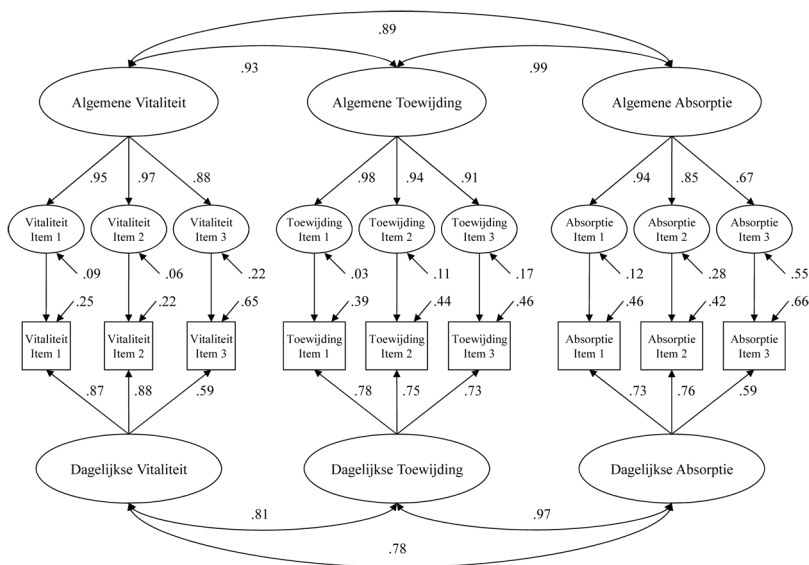
Vervolgens is de multiniveau confirmatieve factoranalyse uitgevoerd. Vaak voeren onderzoekers een factoranalyse uit op de totale covariantiematrix of op de tussenpersoonsmatrix. Aangezien in beide gevallen geen rekening wordt gehouden met de verschillende niveaus en de onafhankelijkheidsassumptie hiermee wordt geschonden, heeft Muthén (1994) de multiniveau confirmatieve factoranalyse (MCFA) ontwikkeld, waarin zowel de binnenpersoons- als tussenpersoonsmatrix wordt gebruikt om de factoranalyse uit te voeren. Figuur 1 illustreert ons MCFA-model met negen geobserveerde indicatoren, die worden aangegeven met rechthoeken. Deze indicatoren zijn de scores op de negen items van de bevoegheidschaal. De geobserveerde variabelen laden allemaal op één van de drie latente factoren (vitaliteit, toewijding en absorptie) op het dagelijkse niveau (onderste gedeelte van het figuur). Op algemeen niveau (het bovenste gedeelte van het figuur) worden de indicatoren aangegeven door middel van een cirkel. Deze zijn niet geobserveerd, maar betreffen gemiddelden van de geobserveerde dagelijkse bevoegdheid. Deze gemiddelden laden op één van de drie geaggregeerde latente variabelen. Het volledige model maakt een associatie tussen beide niveaus. Hierbij wordt er rekening mee gehouden dat de geobserveerde scores een functie zijn van algemene bevoegdheid en dagelijkse bevoegdheid.

De BIC-waarden in tabel 2 laten zien dat het multiniveau 3-factorenmodel (30686.5) beter past op de data dan het multiniveau 1-factormodel (31117.6). De CFI (.96) en SRMR (.04; .04) waarden indiceren een goede fit voor het 3-factorenmodel en de RMSEA (.06) waarde suggereert een acceptabele fit. Voor het 1-factormodel toont alleen de SRMR (.05; .05) een goede fit en de CFI (.90) een acceptabele fit. Om te toetsen of de factorladingen op beide niveaus gelijk zijn, zijn er enkele additionele modellen getest. Ten eerste werd een model getoetst waarbij alle factorladingen gedwongen werden gelijk te zijn op tussen- en binnenpersoonsniveau. Dit model werd vervolgens vergeleken met het vrij geschatte 3-factorenmodel. Het model waarbij de factorladingen op beide niveaus gedwongen waren gelijk te zijn, toonde een significante verhoging van de χ^2 -waarde ($\Delta\chi^2(6) = 47.69$, $p < .001$), wat aangeeft dat niet alle factorladingen gelijk zijn op beide niveaus. Om te toetsen welke factorenloadingen niet gelijk zijn op beide niveaus, werden er additionele analyses gedaan.

In het volgende model dat we hebben getoetst, werden alle factoren gedwongen gelijk te zijn op beide niveaus met uitzondering van item 3 van de vitaliteitschaal. Zoals eerder is aangegeven, is item 3 van de vitaliteitsschaal het enige item dat verwijst naar de ochtend, waardoor dit item niet per se de vitaliteit reflecteert die is ervaren gedurende de dag. Hierdoor zou dit item een betere reflectie kunnen zijn van vitaliteit op tussen- dan op binnenpersoonsniveau. Wanneer dit model werd vergeleken met het vrij geschatte model, resulteerde dit niet in een significante verbetering van de fit ($\Delta\chi^2(5) = 7.35$, *n.s.*). Dit betekent dat de factorladingen van alle items gelijk zijn op beide niveaus behalve vitaliteit item 3, die een betere reflectie is van vitaliteit op tussenpersoonsniveau.

Figuur 1 toont aan dat de factorladingen op tussenpersoonsniveau over het algemeen hoger zijn (.67-.98) dan op binnenpersoonsniveau (.59-.88). De correlaties tussen de drie factoren zijn ook hoger op het tussenpersoonsniveau (.89-.99) dan op het binnenpersoonsniveau (.78-.97). Hoewel de discriminante validiteit op

Figuur 1 Grafische weergave van het resulterende 3-factorenmodel



Tabel 2 Fitindices van het 3-factorenmodel en het 1-factormodel, N = 271

	χ^2	df	CFI	RMSEA	BIC	SRMR
Model						
1-factormodel						
Multiniveau	791.8	54	.90	.10	31117.6	W=.05, B=.05
3-factorenmodel						
Multiniveau	317.4	48	.96	.06	30686.5	W=.04, B=.04

W = binnenpersoonsniveau van het model, B = tussenpersoonsniveau van het model.

beide niveaus vrij laag is, duiden deze correlaties erop dat de discriminante validiteit iets beter is voor de subdimensies van bevoegenheid op het binnenpersoonsniveau.

5 Discussie

5.1 De belangrijkste bevindingen

Het doel van dit onderzoek was het toetsen van de factoriële validiteit van de aangepaste UBES om dagelijkse bevoegenheid te meten. Hiermee wilden we laten zien dat het belangrijk is om de psychometrische eigenschappen van instrumenten die gebruikt worden in dagboekonderzoek te onderzoeken. Het toetsen van de factoriële validiteit van constructen in dagboekonderzoek is van belang omdat het aangeeft of we meten wat we daadwerkelijk beogen te meten. In het huidige onderzoek bleek dat het multiniveau 3-factorenmodel een goede fit met de data had,

wat aangeeft dat het niet nodig is om een conceptueel ander meetinstrument te ontwikkelen om dagelijkse bevoegenheid te meten. Wel bleek dat item 3 van de vitaliteitsschaal een betere reflectie was van bevoegenheid op het tussenpersoonsniveau (algemene bevoegenheid) dan op het binnenpersoonsniveau (dagelijkse bevoegenheid). Voor vervolgonderzoek is het aan te raden om een ander item uit de 17-itemversie van de UBES (Schaufeli & Bakker, 2003; Schaufeli et al., 2002) te kiezen dat wel naar de dag verwijst en niet alleen naar de ochtend (bijv. 'Toen ik vandaag aan het werk was, had ik veel mentale veerkracht').

De drie dimensies van bevoegenheid (vitaliteit, toewijding en absorptie) bleken sterk positief aan elkaar gerelateerd, vooral op het tussenpersoonsniveau. Hiermee wordt ondersteund dat er één algemene factor is, bevoegenheid, die uit drie verschillende factoren bestaat. De discriminante validiteit van deze drie factoren is beter op het binnenpersoonsniveau dan op het tussenpersoonsniveau. We raden dan ook aan om de drie verschillende factoren alleen te gebruiken wanneer er op basis van de theorie verschillende relaties worden verwacht met voorspelers en uitkomsten. Wanneer er verschillende relaties worden verwacht en de totale schaal wordt gebruikt, kan er belangrijke informatie verloren gaan.

5.2 Toekomstig onderzoek

De resultaten van dit onderzoek laten zien dat de kanttekeningen die Sonnentag en collega's (2010) plaatsten bij het aanpassen van algemene vragenlijsten om dagelijkse variabelen te meten, belangrijk zijn bij het meten van dagelijkse bevoegenheid. De factorstructuur blijft overeind, maar een van de items kan in toekomstig onderzoek beter vervangen worden door een ander item. Dit onderzoek geeft enkele handvatten voor toekomstig onderzoek naar de psychometrische eigenschappen van meetinstrumenten in dagboekonderzoek. Wanneer men dagboekonderzoek doet, is het belangrijk om te kijken of de factorstructuur van de gemeten variabelen elke dag hetzelfde is. Met andere woorden, meten we elke dag hetzelfde en meten we wat we beogen te meten? Dit is belangrijk om verschillen tussen gemiddelden over de tijd daadwerkelijk goed te kunnen interpreteren. In longitudinaal onderzoek wordt dit al gedaan (zie bijv. Makikangas, Hatinen, Kinunen & Pekkonen, 2011), maar voor zover wij weten, is dit nog niet gedaan in dagboekonderzoek.

Toekomstig onderzoek naar de psychometrische eigenschappen van meetinstrumenten in dagboekonderzoek doet er ook goed aan om te kijken naar nog specifiekere metingen van dagelijkse variabelen. Hoe zit het bijvoorbeeld met de factorstructuur wanneer taakbevoegenheid wordt gemeten? Taakbevoegenheid is bevoegenheid die afhankelijk is van de taak die een werknemer uitvoert. Het is goed denkbaar dat de bevoegenheid van bijvoorbeeld een verkoper hoger is tijdens het contact met klanten dan tijdens het bijhouden van de administratie. Bevoegenheid zal dus naar verwachting ook over de dag variëren, afhankelijk van de taken die uitgevoerd worden. Het werken met meerdere meetmomenten op één dag kan resulteren in een conceptueel ander meetinstrument, omdat het dan ook andere gevoelens en gedachten kan meten die een lage frequentie hebben of specifiek zijn voor een bepaalde taak, wat kan resulteren in meer geëxtraheerde factoren in een factoranalyse.

Het is voor toekomstig onderzoek ook belangrijk om modellen van algemene en dagelijkse variabelen te integreren en/of aan te passen. Op het gebied van bevologenheid is er op dit moment weinig integratie, terwijl er aanwijzingen zijn dat dit wel mogelijk is. Gebleken is bijvoorbeeld dat zowel algemene als dagelijkse bevologenheid veelal dezelfde voorspellers en uitkomsten heeft. Sommige voorspellers werken echter niet hetzelfde voor algemene en dagelijkse bevologenheid. Onderzoek heeft namelijk laten zien dat werkdruk een negatief verband heeft met algemene bevologenheid, maar positief is gerelateerd aan dagelijkse bevologenheid (Xanthopoulou & Bakker, in press). Een verklaring zou kunnen zijn dat werkdruk op korte termijn zorgt voor het behalen van doelen en bijdraagt aan bevologenheid, terwijl het op lange termijn een stressor wordt.

Een multiniveau confirmatieve factoranalyse heeft veel voordelen, maar er zijn ook enkele kanttekeningen. Een punt van kritiek is dat de algemene bevologenheid die is gebruikt in dit onderzoek, dagelijkse bevologenheid die is geaggregeerd over vijf dagen, geen 'echte' algemene bevologenheid is. Maar deze operationalisatie van algemene bevologenheid is wellicht een betere reflectie van langdurige bevologenheid dan bevologenheid die is gemeten op één punt in de tijd. Sonnentag en collega's (2010) beargumenteerden dat dagelijkse bevologenheid een levendige ervaring reflecteert, terwijl algemene bevologenheid wellicht meer een attitude reflecteert. In het laatste geval wordt mensen namelijk gevraagd om over een langere periode terug te denken, wat het risico op een geheugenbias vergroot. Wanneer bevologenheid geaggregeerd wordt over vijf dagen, zal het meer een levendige ervaring reflecteren dan een attitude. Wanneer de bevologenheidsscores gemiddeld worden over vijf dagen, vermindert dit bovendien de meetfouten (Shiffman, 2007).

Een ander kritiekpunt is dat er geen vuistregel is over het aantal observaties dat nodig is om deze analyse uit te voeren. Eén onderzoek, zoals beschreven in Mok (1995), geeft aan dat er 800 observaties nodig zijn om een multiniveau confirmatieve factoranalyse uit te voeren; het huidige onderzoek bevatte meer dan 800 observaties. Dit wil zeggen dat het kritiekpunt met betrekking tot het aantal benodigde observaties geen rol lijkt te spelen in dit onderzoek. Een sterk punt van het onderzoek is het gebruik van een innovatieve statistische methode om dagelijkse dagboekdata te analyseren. We raden onderzoekers aan om de methode uit dit onderzoek toe te passen om de validiteit te toetsen van aangepaste algemene vragenlijsten om dagelijkse variabelen te meten.

Praktijkbox

- Door bevologenheid en de voorspellers ervan op dagelijkse basis te meten kunnen handvatten worden aangereikt over de manier waarop bevologenheid op korte termijn kan worden gestimuleerd. Dit kan belangrijk zijn op dagen dat een team bijvoorbeeld een belangrijke deadline of vergadering heeft.
- Wanneer bevologenheid op dagelijkse basis wordt gemeten, kan bevologenheid objectiever worden gemeten, omdat medewerkers dan over

een minder lange periode hoeven terug te denken. Dit kan mogelijk resulteren in effectievere interventies om de bevlogenheid te bevorderen.

- Dagboekonderzoek zit dicht op het proces en kan daardoor specifiekere meten wat er nu precies voor zorgt dat iemand meer of minder bevlogen is. Hierdoor komen we niet alleen dichterbij de waarheid, maar kunnen we ook laten zien waarom medewerkers soms meer of minder bevlogen zijn.

Literatuur

- Allen, B.P. & Potkay, C.R. (1981). On the arbitrary distinction between states and traits. *Journal of Personality and Social Psychology*, 41, 916-928.
- Bakker, A.B. (2009). Een overzicht van tien jaar onderzoek naar bevlogenheid. *Gedrag & Organisatie*, 22, 336-353.
- Bakker, A.B. & Bal, P.M. (2010). Weekly work engagement and performance: A study among starting teachers. *Journal of Occupational and Organizational Psychology*, 83, 189-206.
- Bakker, A.B. & Daniels, K. (Eds.) (in press). *A day in the life of a happy worker*. Hove, Sussex: Psychology Press.
- Bolger, N., Davis, A. & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, 54, 579-616.
- Breevaart, K., Bakker, A.B., Demerouti, E. & Hetland, J. (2012). The measurement of state work engagement: A multilevel factor analytic study. *European Journal of Psychological Assessment*, 28, 305-312.
- Dyer, N.G., Hanges, P.J. & Hall, R.J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *Leadership Quarterly*, 16, 149-167.
- James, L.R. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology*, 67, 219-229.
- Makikangas, A., Hatinen, M., Kinnunen, U. & Pekkonen, M. (2011). Longitudinal factorial invariance of the Maslach Burnout Inventory-General Survey among employees with job-related psychological health problems. *Stress & Health*, 27, 347-352.
- Mok, M. (Ed.) (1995). *Sample size requirements for 2-level designs in educational research*. Sydney, Australia: Macquarie University.
- Muthén, B.O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research*, 22, 376-398.
- Muthén, L.K. & Muthén, B.O. (2002). *Mplus VERSION 2.12: Addendum to the Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- Reis, H.T. & Gable, S.L. (Eds.) (2000). Event-sampling and other methods for studying everyday experience. In H.T. Reis & C.M. Judd (Eds.), *Handbook of research methods in social and personality psychology*. Cambridge: Cambridge University Press.
- Schaufeli, W.B. & Bakker, A.B. (2003). *Utrecht Work Engagement Scale: Preliminary Manual*. Department of Psychology, Utrecht University, The Netherlands (available from www.schaufeli.com).

- Schaufeli, W.B. & Bakker, A.B. (2010). Defining and measuring work engagement: Bringing clarity to the concept. In M.P. Leiter & A.B. Bakker (Eds.), *Work engagement: A handbook of essential theory and research* (pp. 10-24). New York: Psychology Press.
- Schaufeli, W.B., Bakker, A.B. & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and Psychological Measurement*, 66, 701-716.
- Schaufeli, W.B., Salanova, M., Gonzalez-Roma, V. & Bakker, A.B. (2002). The measurement of burnout and engagement: A confirmatory factor analytic approach. *Journal of Happiness Studies*, 3, 71-92.
- Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Schweizer, K. (2010). Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment*, 26, 1-2.
- Shiffman, S. (2007). Designing protocols for ecological momentary assessment. In A.A. Stone, S. Shiffman, A.A. Atienza & L. Nebeling (Eds.), *The sciences of real-time data capture: Self-reports in health research* (pp. 27-53). New York: Oxford University Press.
- Sonnentag, S. (2003). Recovery, work engagement, and proactive behavior: A new look at the interface between non-work and work. *Journal of Applied Psychology*, 88, 518-528.
- Sonnentag, S., Dormann, C. & Demerouti, E. (2010). Not all days are equal: The concept of SWE. In M.P. Leiter & A.B. Bakker (Eds.), *Work engagement: A handbook of essential theory and research* (pp. 25-30). New York: Psychology Press.
- Visser, P.S., Krosnick, J.A. & Lavrakas, P.J. (2000). Survey research. In H.T. Reis & C.M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 223-252). Cambridge: Cambridge University Press.
- Wheeler, L. & Reis, H.T. (1991). Self-recording of everyday events: Origins, types, and uses. *Journal of Personality*, 59, 339-354.
- Xanthopoulou, D. & Bakker, A.B. (in press). State work engagement: The significance of within-person fluctuations. In A.B. Bakker & K. Daniels (Eds.), *A day in the life of a happy worker*. Hove, Sussex: Psychology Press.
- Xanthopoulou, D., Bakker, A.B., Demerouti, E. & Schaufeli, W.B. (2009a). Reciprocal relationships between job resources, personal resources, and work engagement. *Journal of Vocational Behavior*, 74, 235-244.
- Xanthopoulou, D., Bakker, A.B., Demerouti, E. & Schaufeli, W.B. (2009b). Work engagement and financial returns: A diary study on the role of job and personal resources. *Journal of Occupational and Organizational Psychology*, 82, 183-200.
- Xanthopoulou, D., Bakker, A.B., Heuven, E., Demerouti, E. & Schaufeli, W.B. (2008). Working in the sky: A diary study on the role of job resources and personal resources. *Journal of Occupational and Organizational Psychology*, 82, 183-200.

The psychometric qualities of measurement instruments in diary studies: an illustration on the basis of the work engagement scale

Kimberley Breevaart, Arnold B. Bakker & Evangelia Demerouti, Gedrag & Organisatie, volume 25, November 2012, nr. 4, pp. 419-432.

Despite the emerging popularity of diary studies, until now little attention has been paid to the psychometric properties of measures used in diary studies. When the time frame of questionnaires in diary studies is adapted, it is important to examine whether the factor structure remains the same. The aim of this article is

to introduce a method to examine the factor structure of daily variables. This method will be illustrated by performing a multilevel confirmative factor analysis on the questionnaire used to measure daily work engagement. The results showed that daily work engagement can be measured by adapting the time frame of the questionnaire to measure enduring work engagement, although it is eligible to replace one of the nine items.

Key words: work engagement, diary studies, measures, psychometric properties, multilevel confirmatory factor analysis