Psychometric assessment of workaholism measures

Cecilie Schou Andreassen
Department of Psychosocial Science, University of Bergen, Bergen, Norway, and The Bergen Clinics Foundation, Bergen, Norway, and
Jørn Hetland and Ståle Pallesen
Department of Psychosocial Science, University of Bergen, Bergen, Norway

Received August 2011
Revised February 2012
Accepted May 2012

Abstract

Purpose – Although the concept of workaholism has existed in the academic literature for decades, exploration of its measurements seems to lag behind. The purpose of this study is to present an investigation of the three most commonly used workaholism measures; the Workaholism Battery (WorkBAT), the Work Addiction Risk Test (WART) and the Dutch Work Addiction Scale (DUWAS) in terms of their cross-validation, their temporal stability and their factor structure.

Design/methodology/approach – The three measures were administered to 661 cross-occupational Norwegian workers. A total of 368 of these completed the same measures 24-30 months later.

Findings – The cross-validation showed that the correlations between the scores of the different instruments were too low to conclude that they measure the same construct. The 24-30 month test-retest reliability coefficients for the measures revealed that the scores were quite stable over time. None of the previously suggested factor solutions for the three measures had a good fit with the data. Explorative factor analyses supported a four-factor solution for the WorkBAT and for the WART. A two-factor solution for the DUWAS was found.

Research limitations/implications – All the data are based on self-report, which might bias the results.

Practical implications – Different workaholism measures cannot be used interchangeably.

Originality/value – This is the first study that cross-validates the three most used workaholism measures and which investigates the reliability of these instruments over a long-term period (24-30 months).

Keywords Assessment, Workaholism, Psychometrics, Scale, Work addiction

Paper type Research paper

Originally, the expression "workaholism" was first introduced in academic literature as an "addiction to work, the compulsion or uncontrollable need to work incessantly" (Oates, 1971). Later, several researchers have suggested further definitions of this concept. Most of these describe workaholism as a chronic pattern of overindulgence in work, long working hours, working more than is demanded by implicit and explicit norms as well as self-absorption in work (Ng *et al.*, 2007; Porter, 1996; Robinson, 1998; Scott *et al.*, 1997; Spence and Robbins, 1992). Still, there is some disagreement about the concept. Some argue for example, that workaholism primarily is a positive attribute or behaviour tendency, encompassing among others high work motivation (Machlowitz, 1980; Scott *et al.*, 1997). Others, on the other hand, define and regard workaholism first and foremost as a negative entity, characterised by compulsiveness and rigidity (Oates, 1971; Robinson, 1998; Schaufeli *et al.*, 2009). The latter view seems to be the prevailing

Journal of Managerial Psychology
Vol. 29 No. 1, 2014
pp. 7-24

0268-3946
DOI 10.1108/JMP-05-2013-0143

today (Taris *et al.*, 2010). One contemporary definition of workaholism is "being overly concerned about work, to be driven by an uncontrollable work motivation, and to spend so much energy and effort into work that it impairs private relationships, spare-time activities and/or health". This definition suggest that workaholism is a subtype of heavy work investment which is defined in terms of both a time dimension and an effort dimension (Snir and Harpaz, 2009).

So far only a few empirical validated measures of workaholism have been developed. The most well-known of these is the workaholism battery (WorkBAT) which was developed by Spence and Robbins (1992). They suggested that a typical workaholic is heavily involved in work, feels motivated to work by an inner drive, and experiences low levels of enjoyment of work. Accordingly, they created three self-report scales reflecting these three concepts. In the scale construction process of the WorkBAT an initial pool of items were first administered to students enrolled in introductory psychology classes at the University of Texas at Austin. Items with poor psychometric properties were excluded or rewritten before the scale was administered to a new set of students. The results indicated good psychometric properties of all items. The revised scales were then administered to an adult sample of about 800 social workers as these were regarded representative of a homogenous group of workers whose work responsibilities were open-ended. Valid responses were received from 291 workers. The final instrument comprised a total of 25 items answered on a five-point Likert scale ranging from "strongly disagree" to "strongly agree" distributed over the three subscales: work involvement (eight items), drive (seven items) and enjoyment of work (ten items). Based on cluster analysis the authors identified several groups of workers. Workaholics scored above the mean on work involvement and drive but below the mean on enjoyment of work, whereas for example work enthusiasts scored above the mean on work involvement and enjoyment of work and below the mean on drive (Spence and Robbins, 1992).

The first measure of workaholism to be developed, however, was the work addiction risk test (WART) (Robinson, 1989). Potential items were drawn from symptoms reported by clinicians in diagnoses of workaholism as they worked with clients and families on the problem of work addiction. The WART comprises 25-items, all rated on a four-point Likert scale ranging from 1 = "never true to" 4 = "always true". Although the scale often is used to create a single composite score (Robinson, 1989; Robinson, 1998), studies by its constructors indicated that its items seemed to be distributed between five factors/subscales: compulsive tendencies (nine items), control (seven items), impaired communication/self-absorption (five items), inability to delegate (one item) and self-worth (two items) (Flowers and Robinson, 2002). Several studies have attested to the basic psychometric properties of the WART in terms of two-week test-retest reliability (Robinson *et al.*, 1992), split-half reliability (Robinson and Post, 1995), face validity (Robinson and Phillips, 1995) as well as convergent and discriminative validity (Robinson, 1999).

More recently Schaufeli *et al.* (2009) developed a new workaholism scale, denoted the Dutch work addiction scale (DUWAS). In the construction process they administered the nine item compulsive tendencies subscale of the WART and the seven items of the drive subscale of the WorkBAT to a large group of workers in The Netherlands and Japan. All items were scored on a four-point Likert scale ranging from 1 = "totally disagree" to 4 = "totally agree". Half of the respondents in each country

were used for scale construction (exploratory sample) and the other half for cross-validation (confirmatory sample). A three factor solution was first found. Items with high cross-loadings and with loadings on the third and smallest factor were removed. A clear two-factor solution (working excessively and working compulsively) then emerged. Later confirmatory analyses showed that the two-factor structure had a good fit with the data. The two new scales also showed good convergent and discriminative validity (Schaufeli *et al.*, 2009). The DUWAS has also shown good psychometric properties in other studies (del Libano *et al.*, 2010). Based on a national database of Dutch workers subjects scoring above certain cut-offs on both subscales can be classified as workaholics (Schaufeli *et al.*, 2011).

Although the three abovementioned scales clearly seem to be the dominating ones and the ones which will be the focus of empirical investigation in the present article, other workaholism related scales have been developed that deserve to be mentioned. One in the workaholism subscale (18 items) of the schedule for non-adaptive and adaptive personality (SNAP) which is a 375 true/false questionnaire related to normal and abnormal personality traits (Clark, 1996). So far the workaholism subscale of the SNAP has mostly been used within different military samples (Moes *et al.*, 1996; Woods *et al.*, 2008). Mudrack and Naugthon (2001) argued that workaholism should be viewed as a tendency to perform non-required work and to control the work of others and constructed accordingly two subscales reflecting these tendencies: non-required work scale which assesses how much time and energy respondents spend on thinking about ways to improve their work and initiate new projects and the control of others subscale assessing fixing problems created by others, checking the work of others, taking the responsibility of the work of others and dealing with crises. A third instrument was developed by Buelens and Poelmans (2004). They based their scale construction on the workaholism triad of Spence and Robbins (1992). By using a total of 20 items from different scales, among others the WorkBAT and the WART, in an explorative factor analysis they found support for a three-factor solution reflecting the three dimensions of Spence and Robbins (1992): enjoyment, work involvement and feeling driven. Based on these dimensions they identified eight clusters of workers (Buelens and Poelmans, 2004).

Although the WorkBAT, the WART and the DUWAS seem to be the most used measures of workaholism, they are still not without weaknesses and unanswered questions. For one there has been suggested that the WorkBAT and the WART not necessarily measure the same construct. The WART has for example been strongly linked to type A behaviour (McMillan *et al.*, 2001), whereas the WorkBAT mainly has been validated against measures of job involvement, health complaints, hours worked, perfectionism and non-delegation (Andreassen *et al.*, 2007, 2010; McMillan *et al.*, 2001). Although there for long has been suggested to cross-validate different workaholism measures against each other (McMillan *et al.*, 2001) and although there have been some crude attempts to do so (Huang *et al.*, 2010) there is still conspicuously little knowledge about how different workaholism measures relate to each other.

Secondly, little is known about the stability of the scores of workaholism. According to the trait perspective, workaholism is regarded as a relatively stable characteristic of the individual (McMillan *et al.*, 2003). Accordingly one would expect measures of workaholism to show high intra-individual stability over time. There has been studies which have investigated the long-term test-retest reliability of some of these scales

(Burke *et al.*, 2002), but we have no knowledge of studies which for example have employed a test-retest time frame which exceeds six months. Thus more knowledge about the long term stability of workaholism is needed in order get a better understanding of the concept.

Third, and lastly, there seem to exist great differences when it comes to the factor structure of the three abovementioned instruments. In terms of the WorkBAT, studies have repeatedly failed to support the three factor structure suggested by Spence and Robbins (1992). In one study support for a reduced two-factor solution was found: enjoyment (item 2, 4, 7, 9, 10, 11, 17) and drive (3, 14, 18, 20, 21, 22, 25) (McMillan *et al.*, 2002). Also Kanai *et al.* (1996) found support for a two-factor solution among Japanese workers: enjoyment of work (item 2, 4, 7, 9, 10, 11, 16, 17, 22, 23) and drive (3, 5, 12, 14, 18, 20, 21, 24) as well as did Ersoy-Kart (2005) with a Turkish version: enjoyment (item, 2, 4, 7, 8, 9, 10, 11, 17, 19) and drive (item 3, 14, 15, 16, 18, 20, 21, 22, 23, 24, 25). In a recent study from Taiwan support for the following five-factor model was found: enjoyment (item 2, 4, 7, 9, 10, 11, 17), work involvement-enjoyment (item 12, 13, 15, 16, 19), drive-work involvement (item 3, 5, 20, 21, 23, 24), drive (item 14, 18, 22), and work involvement (item 1, 6, 8) (Huang *et al.*, 2010).

Non-convergent findings concerning its factor structure also seems to be the case for the WART. Albeit Robinson (1989) argues for a one factor solution (by adding the sum of all items to make one overall score), data have offered support for several other factor solutions. Support for the following five-factor solution was reported by Robinson and Post (1994): overdoing (item 3, 5, 6, 7, 15), self-worth (item 8, 9, 10, 19, 20), control-perfection (item 1, 2, 4, 11, 12, 16, 17, 18, 21, 22, 25), intimacy (item 23, 24), and future reference/mental preoccupation (item 13, 14). Flowers and Robinson (2002) later found support for another five factor solution of the WART: compulsive tendencies (item 3, 5, 6, 7, 8, 15, 18, 19, 20), control (item 2, 4, 11, 12, 16, 17, 22), impaired communication/self-absorption (item 13, 21, 23, 24, 25), self-worth (item 9, 10), and inability to delegate (item 1). Previously they had also argued for a model comprising one second-order factor reflecting the following first order factors: compulsive tendencies (item 3, 5, 6, 7, 8, 15, 18, 19, 20), control (item 2, 4, 11, 12, 16, 17, 22), impaired communication/self-absorption (item 13, 21, 23, 24, 25), and self-worth (item 9, 10) (Robinson, 2001).

Concerning the DUWAS we have no knowledge of its factor structure being investigated by other researchers than its constructors. Thus, its proposed two-factor solution should be investigated by more studies.

Against this backdrop we decided to:

(1) Investigate the relationship between the original three dimensions of the WorkBAT (enjoyment of work, drive and work involvement) with the composite score of the WART as well as with the five dimensions of the WART suggested by its constructors (compulsive tendencies, control, impaired communication/self-absorption, self-worth, and inability to delegate) and with the two dimensions of the DUWAS (working excessively and working compulsively).

(2) To investigate the test-retest reliability and stability of the three abovementioned instruments over a 24-30 month period by calculating intraclass correlation coefficients. Such correlation coefficients have the advantage over other correlations coefficients (e.g. Pearson's product-moment

correlation coefficient) that it can detect the existence of systematic changes in scores over time. If for example all the scores are doubled (systematic changes) from test to retest then the Pearson's product-moment correlation coefficient would indicate a perfect test-retest reliability, whereas the intraclass correlation coefficient would be considerably lower due to the systematic change over time (Yen and Lo, 2002).

(3) To investigate the fit with all the above-mentioned factor structures for the three measures using confirmatory factor analyses and in the case of poor fit with all models, to investigate the factor structure of the instrument by using explorative factor analysis.

Methods

Samples and procedures

We administered internet-based questionnaires to 1,300 Norwegian cross-occupational employees from seven different organisations (managers of a major national pharmaceutical company ($n = 127$), employees of a regional healthcare sector company ($n = 96$), a national TV station ($n = 172$), two different human resource (HR) consultancy companies ($n = 80$), and employees from two university faculties ($n = 186$)) who were recruited to participate in a study of their psychosocial working environment. A total of 661 completed the questionnaires, yielding a response rate of 51 per cent, ranging from 39 per cent (university) to 68 per cent (TV) within the different companies. This final sample consisted of 360 (54 per cent) females and 301 (46 per cent) males, whose ages ranged from 16 to 72 years (mean age = 42.6, SD = 10.5). The majority of the respondents were married or living with a partner (67 per cent), were living with children (53 per cent), and had education at university level (87 per cent). The period of service in the companies ranged from new employees to employees who had tenure longer than 20 years. Most of the employees worked full time (88 per cent) and had worked in the organisations for between 0 and 10 years (86 per cent).

Between 24 and 30 months following the administration of the abovementioned survey, we re-administrated a survey to the 661 participants of wave 1. A total of 368 (mean age = 46.4, SD = 10.1), 175 males and 193 females, took part in the second wave, yielding a response rate of 55.7 per cent, The sample in wave 2 comprised leaders of a major pharmaceutical company ($n = 63$), employees from a regional healthcare sector company ($n = 57$), a national TV station ($n = 87$), two different HR consultancy companies ($n = 41$), and two university faculties ($n = 120$). As their e-mail addresses comprised one of the variables in each of the data sets, the data sets were merged so that responses to wave 1 could be aligned with responses from wave 2 for each single individual. The Regional Committee for Medical and Health Research Ethics in Western Norway had approved the study.

Instruments

Demographics. Questions about age and gender were asked. We also asked respondents to provide an estimate of the total number of hours worked every week including overtime, secondary jobs, and other paid and unpaid jobs except for domestic chores.

The work addiction risk test (WART). The WART scale comprises 25 items. Each item is answered on a four-point Likert scale ranging from "never true" (1) to "always true" (4) (Robinson, 1989).

The workaholism battery (WorkBAT). The WorkBAT comprises 25 items originally distributed along three subscales: work involvement (item 1, 6, 8, 12, 13, 15, 21, 24), drive (item 3, 5, 14, 18, 20, 22, 25) and enjoyment of work (item 2, 4, 7, 9, 10, 11, 16, 17, 19, 23). Each item is rated on a five-point Likert scale (Spence and Robbins, 1992) ranging from "strongly agree" (5) to strongly disagree (1).

The Dutch work addiction scale (DUWAS). The DUWAS comprises ten items, five assessing working excessively (all taken from the WART) and five assessing working compulsively (four items taken from the WorkBAT and one from the WART). Each item is scored along a four-point scale ranging from 1 = "totally disagree" to 4 = "totally agree" (Schaufeli *et al.*, 2009). In the present study the DUWAS items were administered as part of the WART and the WorkBAT. Hence, the response alternatives deviated somewhat from the original DUWAS. In order to transform the five-point responses alternatives of the WorkBAT items to a scale ranging from 1 to 4 the following recoding took place: 1 = 1, 2 = 1.75, 3 = 2.5, 4 = 3.25 and 5 = 4, respectively.

Statistics

In order to cross-validate the three workaholism instruments Pearson product-moment correlation coefficients between their subscales were calculated. Long-term stability and test-retest reliability of the scale scores was calculated by intraclass correlation coefficients (ICC). These analyses were conducted with SPSS, version 19. The fit of the different proposed factor solutions for the three instruments were investigated by confirmatory factor analyses using AMOS, version 18.0. The RMSEA, the CFI and the TLI were used as fit indexes. As a rule of thumb, for a model with acceptable fit to the data, these indexes should be <0.08, >0.90 and >0.90 respectively, whereas the corresponding values for a good fit would be <0.06, >0.95, and >0.95 respectively (Hu and Bentler, 1999). All error terms were expected to be uncorrelated, however correlations were assumed between all the latent variables. In the case of poor fit, the items of the instruments in question would be subjected to an explorative factor analysis based upon a principled components analysis. Parallel analysis would be used to determine the number of factors to be retained (Horn, 1965). In the case of more than one factor being retained, direct oblim rotation would be used, as this allows for correlation between the rotated factors. The explorative factor analyses would be carried out with SPSS, version 19.0. Only factor loadings ≥ 0.40 would be considered as substantial.

Results

Table I shows the Pearson moment-product correlation coefficients between the different workaholism scales, included their subscales. The work involvement subscale of the WorkBAT showed rather low correlations with the other scales, the highest being 0.39 with the compulsive tendencies subscale of the WART. The drive scale of the WorkBAT showed rather high correlations ($r \geq 0.60$) with the compulsive tendencies subscale of the WART and naturally with the working compulsively subscale of the DUWAS. Enjoyment of work of the WorkBAT had overall the lowest correlations with other scales, the highest being 0.21 with the working excessively

Table I. Means, standard deviations, Cronbach's alphas and Pearson moment-product correlation coefficients between the different workaholism measures

	Mean	SD	α	2	3	4	5	6	7	8	9	10	11
Workaholism battery													
1. Work involvement	23.12	4.81	0.63	0.30**	0.14**	0.29**	0.39**	0.11**	0.15**	0.10**	0.12**	0.33**	0.32**
2. Drive	21.82	5.84	0.82		0.03	0.58**	0.61**	0.39**	0.39**	0.10*	0.23**	0.47**	0.95**
3. Enjoyment of work	33.89	6.36	0.84			0.19**	0.19**	0.10**	0.18**	0.06	0.15**	0.21**	0.06**
Work addiction risk test													
4. Total composite score	57.15	7.36	0.86				0.86**	0.81**	0.73**	0.29**	0.43**	0.73**	0.59**
5. Compulsive tendencies	22.83	3.40	0.77					0.52**	0.50**	0.20**	0.31**	0.89**	0.64*
6. Control	14.69	2.78	0.75						0.45**	0.19**	0.28**	0.41**	0.39**
7. Impaired communication	9.52	1.96	0.59							0.15**	0.23**	0.41**	0.39**
8. Inability to delegate	2.64	0.54									0.06	0.19**	0.10*
9. Self-worth	5.51	0.89	0.36									0.28**	0.22**
Dutch work addiction scale													
10. Working excessively	13.73	2.06	0.69										0.47**
11. Working compulsively	12.40	3.10	0.63										

Note: n=661

subscale of the DUWAS. The composite score of the WART had naturally high correlations with most of its subscales, and correlated highly with the two DUWAS subscales.

The compulsive tendencies subscale of the WART correlated highly with the drive subscale of the WorkBAT as well as with the two subscales of the DUWAS. The control subscale and the impaired communication subscale of the WART had moderate correlations with most of the other scales. The inability to delegate subscale and the self-worth subscale of the WART had overall relatively low correlations with the other scales. The two subscales of DUWAS had overall moderate to high correlations with each other as well as with other workaholism scales.

Table II shows the ICCs reflecting the 24-30 month test-retest reliability of the different workaholism measures. Except for inability to delegate which had an ICC of 0.32, the ICC's for the other workaholism measures ranged between 0.56 (impaired communication and self-worth subscales of the WART) and 0.70 (composite score of the WART).

Table III shows the results of the confirmatory factor analyses for different factor structures that previously have been suggested for the three workaholism measures. None of the models showed a good fit on any parameter. For the WorkBAT the two-factor model of McMillan *et al.* (2002) had an acceptable fit with the data according to the RMSEA and the CFI.

The explorative factor analysis of the WorkBAT items supported a four-factor solution according to the parallel analysis (the five highest Eigenvalues were 4.54, 3.49, 2.03, 1.30, and 1.17, respectively, whereas the five highest randomly generated Eigenvalues were 1.37, 1.32, 1.27, 1.24 and 1.20, respectively). None of the items had cross-loadings on other factors. Item 12 and item 19 did not load on any factor. The results are presented in Table IV. The explorative factor analysis of the WART items also supported a four-factor solution according to the parallel analysis (the highest Eigenvalues were 5.80, 1.83, 1.56, 1.31, and 1.18, respectively). None of the items had

	ICC
Workaholism battery	
1. Work involvement	0.65
2. Drive	0.64
3. Enjoyment of work	0.61
Work addiction risk test	
4. Total composite score	0.70
5. Compulsive tendencies	0.63
6. Control	0.69
7. Impaired communication	0.56
8. Inability to delegate	0.32
9. Self-worth	0.56
Dutch work addiction scale	
10. Working excessively	0.61
11. Working compulsively	0.65

Note: $n = 340$

Table II.
Intra-class correlation coefficients (ICC) for the 24-30 month test-retest reliability of the different workaholic measures

Table III.
Fit indexes for different factor models of workaholism measures

	χ^2	df	*p*	RMSEA (90% CI)	CFI	TLI
Workaholism battery						
Spence and Robbins (1992) three-factor model	1,165.0	272	0.00	0.071 (0.066-0.075)	0.805	0.785
Kanai *et al.* (1996) two-factor model	867.9	134	0.00	0.091 (0.085-0.097)	0.778	0.747
McMillan *et al.* (2002) two-factor model	348.8	76	0.00	0.074 (0.066-0.082)	0.904	0.885
Ersoy-Kart (2005) two-factor model	1,129.6	169	0.00	0.093 (0.088-0.098)	0.741	0.709
Huang *et al.* (2010) five-factor model	1,345.9	242	0.00	0.083 (0.079-0.087)	0.686	0.725
Work addiction risk test						
Robinson (1989) one-factor model	1,365.3	275	0.00	0.078 (0.073-0.082)	0.690	0.662
Robinson and Post (1994) five-factor model	993.9	265	0.00	0.065 (0.060-0.069)	0.793	0.765
Robinson *et al.* (2001) second-order factor model	964.8	226	0.00	0.070 (0.066-0.075)	0.778	0.751
Flowers and Robinson (2002) five-factor model	997.4	243	0.00	0.069 (0.064-0.073)	0.776	0.746
Dutch work addiction scale						
Schaufeli *et al.* (2009) two-factor model	281.1	34	0.00	0.105 (0.094-0.116)	0.849	0.801

Item	Wording	Original scale[c]	Factor loadings[b] I	II	III	IV
1	When I have free time I like to relax and do nothing serious	WI			0.69	
24	I get bored and restless on vacations when I haven't anything productive to do[a]	WI			−0.43	
13	I spend my free time on projects and other activities	WI				0.77
21	Between my job and other activities I'm involved in I don't have much free time	WI				0.68
15	I like to use my time constructively, both on and off the job	WI				0.63
6	I like to relax and enjoy myself as often as possible[a]	WI			0.70	
8	I really look forward to the weekend – all fun, no work[a]	WI			0.64	
12	Wasting time is as bad as wasting money	WI				
25	I seem to have an inner compulsion to work hard	D		0.76		
22	I often feel there is something inside me that drives me to work hard	D		0.75		
18	It is important to me to work hard, even when I don't enjoy what I'm doing	D		0.72		
20	I often find myself thinking about work, even when I want to get away from it for a while	D		0.62		
3	I feel guilty when I take time off work	D		0.57		
14	I feel obliged to work hard even when it is not enjoyable	D		0.79		
5	I often with I weren't so committed to my work	D		0.50		
7	My job is so interesting that it often doesn't seem like work	WE	0.75			
19	When I get involved in an interesting project it's hard to describe how exhilarated I feel	WE				
16	I lose track of time when I'm involved in a project	WE	0.54			
9	I do more work than is expected of me strictly for the fun of it	WE	0.71			
10	Most of the time my work I very pleasurable	WE	0.62			
23	Sometimes I enjoy work so much I have a hard time stopping	WE	0.76			
2	I like my work more than most people do	WE	0.72			
4	My job is more like fun than work	WE	0.77			
11	I seldom find anything to enjoy about work[a]	WE	−0.55			
17	Sometimes when I get up in the morning I can hardly wait to get to work	WE	0.61			

Note: $n = 661$; [a]Originally reversed items, [b]Only loadings ≥ 0.40 are shown; [c]WI = work involvement, D = drive, WE = work enjoyment

Table IV.
Pattern matrix of the explorative factor analysis of the workaholism battery items

cross-loadings on other factors. Item 1, 4, 8, 18, 19 and 21 did not load on any factor. The results are presented in Table V. The results for DUWAS supported a two-factor solution (the three highest Eigenvalues were 3.59, 1.40, and 1.03, respectively, whereas the three highest randomly generated Eigenvalues were 1.19, 1.13, and 1.09, respectively). All items loaded on one of the two factors only. The results are presented in Table VI. The requirements for all three explorative factor analyses in terms of the

Item	Wording	Original scale[b]	Factor loadings[a] I	II	III	IV
1	I prefer to do most things myself rather than ask for help	ID				
2	I get very impatient when I have to wait for someone else or when something takes too long, such as long-slow moving lines	C				0.49
3	I seem to be in a hurry and racing against the clock	CT	0.65			
4	I get irritated when I am interrupted while I am in the middle of something	C				
5	I stay busy and keep many irons in the fire	CT	0.78			
6	I find myself doing two or three things at one time, such as eating lunch and writing a memo, while talking on the telephone	CT	0.75			
7	I overly commit myself by biting off more than I can chew	CT	0.46			
8	I feel guilty when I am not working on something	CT				
9	It is important that I see the concrete results of what I do	SW				0.54
10	I am more interested in the final result of my work than in the process	SW				0.70
11	Things just never seem to move fast enough or get done fast enough for me	C				0.53
12	I lose my temper when things don't go my way or work out to suit me	C		0.69		
13	I ask the same question over again, without realizing it, after I've already been given the answer once	IC/SA		0.63		
14	I spend a lot of time mentally planning and thinking about future events while tuning out the here and now			0.56		
15	I find myself still working after my co-workers have called it quits	CT				
16	I get angry when people don't meet my standards of perfection	C		0.51		
17	I get upset when I am in situations where I cannot be in control	C		0.70		
18	I tend to put myself under pressure with self-imposed deadlines when I work	CT				
19	It is hard for me to relax when I'm not working	CT				
20	I spend more time working than on socializing with friends, on hobbies or on leisure activities	CT			0.63	
21	I dive into projects to get a head start before all the phases have been finalized	IC/SA				
22	I get upset with myself for making even the smallest mistake	C		0.55		
23	I put more thought, time, and energy into my work than I do into my relationships, with my spouse, (or lover) and family	IC/SA			0.78	
24	I forget, ignore or minimize important family celebrations such as birthdays, reunions, anniversaries or holidays	IC/SA			0.75	
25	I make important decisions before I have all the facts and have a chance to think them through thoroughly	IC/SA			0.46	

Note: $n = 661$; [a]Only loadings ≥ 0.40 are shown; [b]ID = inability to delegate, C = control, CT = compulsive tendencies, SW = self worth, IC/SA = impaired communication/self absorption

Table V.
Pattern matrix of the explorative factor analysis of the work addiction risk test items

Kaiser-Meyer Olkin measure of sampling adequacy (>0.60) and the Bartlett's test of sphericity ($p < 0.05$) were fulfilled. Taken together the explorative factor analyses supported a four-factor solution for the WorkBAT and for the WART whereas a two-factor solution for the DUWAS was supported by the data.

Discussion

Implications for theory

We start this discussion by commenting on the correlation coefficients between the different workaholism measures. Concerning the WorkBAT, the work involvement and the enjoyment of work subscales had overall low to moderate correlations with the WART and its subscales and with the two subscales of the DUWAS, clearly indicating that they do not measure the same construct. The drive subscale of the WorkBAT correlated however somewhat higher with some of the other scales, such as the total composite score of the WART (0.58) and the compulsive tendencies subscale of the WART (0.61). It naturally correlated highly (0.95) with the working compulsively subscale of the DUWAS as these two subscales have four items in common. However, if the WorkBAT measured the exact same construct as the WART and the working excessively subscale of the DUWAS, we should expect to see far higher correlation coefficients than we actually did. Hence, the WorkBAT did not demonstrate adequate convergent validity with other workaholism scales. The same goes for the WART. The composite score of the WART and the compulsive tendencies subscale correlated fairly highly with the two DUWAS subscales, but this probably reflects their shared items. Overall, the convergent validity of the three workaholism measures does not seem adequate (Pedazur and Schmelkin, 1991). Still, it might be argued that the three

Table VI.
Pattern matrix of the explorative factor analysis of the Dutch work addiction scale items

Item	Wording	Original scale[b]	Factor loadings[a] I	II
1	I seem to be in a hurry and racing against the clock	WE		0.73
2	I find myself still working after my co-workers have called it quits	WE		0.44
3	I stay busy and keep many irons in the fire	WE		0.82
4	I spend more time working than on socializing with friends, on hobbies or on leisure activities	WE		0.41
5	I find myself doing two or three things at one time, such as eating lunch and writing a memo, while talking on the telephone	WE		0.76
6	It is important to me to work hard, even when I do not enjoy what I am doing	WC	0.77	
7	I often feel there is something inside me that drives me to work hard	WC	0.68	
8	I feel obliged to work hard, even when it is not enjoyable	WC	0.85	
9	I feel guilty when I take time off work	WC	0.70	
10	It is hard for me to relax when I am not working	WC	0.46	

Note: $n = 661$; [a]Only loadings ≥ 0.40 are shown; [b]WE = working excessively, WC = working compulsively

instruments may converge better when it comes to identifying subjects as workaholics. Cut-offs have been suggested, for example, for the WART (Robinson, 1998) and the DUWAS (Schaufeli *et al.*, 2011). It has also been suggested to combine scores of the three WorkBAT scales to classify respondents as workaholics, although no specific cut-offs have been suggested in this regard (Spence and Robbins, 1992). We did not investigate the convergence between the measures in terms of categorisation of respondents, as most researchers still seem to use the scores of the three workaholism instruments as continuous variables, not creating categories. Thus, the lack of convergent validity is a problem. This problem seems to be rooted in the fact that there is not yet any consensus of how to define and understand workaholism (Burke, 2006). The construction process of the different workaholism measures also seem to diverge, especially in terms of their theoretical foundation and methodological approaches (Robinson, 1989; Schaufeli *et al.*, 2009; Spence and Robbins, 1992).

When it comes to the 24-30 month test-retest reliability of the different workaholism measures, this was investigated by the intraclass correlation coefficient (ICC). This takes into consideration not only the relative stability, but also the absolute stability of the measure; hence it expresses the test-retest reliability in terms of both unsystematic and systematic error (Weir, 2005). Overall, the results indicated quite a high test-retest reliability of the three measures. Except for the inability to delegate subscale of the WART, which comprises only one item and therefore naturally has a rather low test-retest reliability, all the scales and subscales had an ICC ranging from 0.56 to 0.70. Most of these values are actually comparable to stability coefficients for a wide range of personality scales for the same test-retest interval (Costa and McCrae, 1994). This may indicate that the construct measured by the workaholism instruments can be regarded as a stable entity or a personality trait (McMillan *et al.*, 2003). However it should be noted in this regard that the workers in the test-retest sample were all employed in the same organisation, and probably kept the same job, at the time of both the test and the retest. Thus, a stable work situation may have contributed to the high test-retest reliability. The latter point is in line with studies showing that situational variables, such as organisational values, may influence work-family imbalance (Burke, 2001). For future studies on the long-term stability of workaholism, it would therefore be of great interest to compare individuals who keep the same job over time with individuals who change job/organisation.

The last point addressed in the present study was the factor structure of the workaholism instruments. We did not find strong support for any of the previous factor solutions for the WorkBAT in our cross-occupational Norwegian sample. The two-factor solution of the WorkBAT proposed by McMillan *et al.* (2002) did, however, have an acceptable fit as indicated by both the RMSEA and CFI. This model did not find support for the work involvement subscale of the WorkBAT and as such is in line with other researchers' pointing to the psychometric problems pertaining to this subscale in particular (Ersoy-Kart, 2005; Kanai *et al.*, 1996). Previous models for the WART were also not supported in the present study (Flowers and Robinson, 2002; Robinson, 1989; Robinson *et al.*, 2001; Robinson and Post, 1994). So far, the factor structure of the DUWAS has been investigated to a limited extent outside The Netherlands, Japan and Spain (del Libano *et al.*, 2010; Schaufeli *et al.*, 2009). In the present Norwegian sample, the fit of the proposed model for DUWAS was not acceptable. Due to the present state of affairs, we conducted explorative factor analyses

with the items of the different workaholism instruments. For the WorkBAT, we found support for a four-factor solution. Two of the factors, I and II, correspond well to the enjoyment of work and the drive subscale, respectively and are thus in line with most previous studies on the factor structure of the WorkBAT (Ersoy-Kart, 2005; Kanai *et al.*, 1996; McMillan *et al.*, 2002). The items, originally belonging to the work involvement subscale, were distributed over two factors, III and IV. Factor III seems to reflect the ability to relax, whereas factor IV comprises items related to being an active and involved person. Our explorative factor analysis of the WorkBAT items confirms the poor psychometric properties of the work involvement subscale (Ersoy-Kart, 2005; Kanai *et al.*, 1996; McMillan *et al.*, 2002). Two items did not load on any factors (12, 19). Also for the WART we found support for a four-factor solution. Factor I seems to contain items pertaining to overwork and has a large overlap (four of five items) with the overdoing factor identified by Robinson and Post (1994). Factor II does seem to correspond to a certain degree with the control-perfection subscale of Robinson and Post (1994) and seems to reflect perfectionism and distractibility. Factor III contains the two items of the intimacy subscale suggested by Robinson and Post (1994) in addition to item 20 and 25. Its content mainly mirrors the work-family interface. The items loading on factor IV are mainly related to impatience. It contains the two items of the self-worth subscale proposed by Robinson *et al.* (2001). A total of seven items of the WART (item 1, 4, 8, 15, 18, 19, 21) did not load on any of the four factors. The present study, as well as previous studies (Flowers and Robinson, 2002; Robinson, 1989; Robinson *et al.*, 2001; Robinson and Post, 1994), indicates a more unclear factor structure of the WART than the WorkBAT.

Although the confirmatory factor analysis for the proposed factor structure of the DUWAS (Schaufeli *et al.*, 2009) showed a poor fit with the data, we still found support for the two-factor structure in the explorative factor analysis. The reason why the factor structure was not supported by the confirmatory factor analysis is most likely related to the fact that several individual items had relatively low loadings.

Implications for managerial practice

The results from the present study suggest that although the different measures of workaholism investigated in the present study seem to be stable over time, the question about their conceptual validity does not seem to be adequately answered. For managerial practice this means that one needs to be careful when interpreting scores from workaholism measures. The results from the present study also demonstrate that one cannot compare results across different workaholism measures without reservations, as they do not seem to have a large enough conceptual overlap (low convergent validity).

Limitations and directions for future research

In terms of strengths and limitations of the present study, it should be noted that it was conducted outside the US and central Europe where most of the workaholism studies have been conducted so far. This may however have influenced the results as culture can have a great impact on questionnaire data, especially when there is lack of substantial equivalence (cultural differences in the phenomena being assessed), lack of functional equivalence (cultural differences in willingness/motivation to report attitudes, feelings and behaviour), lack of conceptual equivalence (cultural differences

in meaning attached to questionnaire items), lack of linguistic equivalence (cultural differences in the meaning conveyed by some words/sentences), as well as lack of scalar equivalence (cultural differences in how scales or response alternatives are perceived) (Lonner and Ibrahim, 1996). A strength of the present study is that the sample was quite large and the respondents were recruited from several different professions. As far as we know, this is the first study which has cross-validated three workaholism instruments and the first study which has investigated the test-retest reliability of workaholism instruments over a time span exceeding six months. One limitation of the present study is that all data were based on self-report, thus, the results may have been influenced by the common method bias (Podsakoff *et al.*, 2003). Several subscales of the workaholism instruments investigated in the present study had rather low Cronbach's alphas, suggesting low internal consistency. This may have impacted the results by reducing the magnitude of relationships between the constructs being investigated. However, it should be kept in mind that many of the subscales in question contained relatively few items, which naturally lowers the Cronbach's alpha (Streiner and Norman, 2008). For the DUWAS we did not use the original response alternatives, and the items of the DUWAS taken from the WorkBAT had to be transformed from a 1-5 range scale to a 1-4 range scale. Analyses (results not reported here) showed that this had only negligible effects on the results.

Overall we conclude that the different workaholism measures correlate too low with one other to reflect the same construct. Although there is obviously a certain overlap, it is too small. Future studies and theoretical developments should therefore aim at establishing consensus about the definition of workaholism, and constructing a corresponding well-validated workaholism scale. Although it may be tempting to create a new workaholism scale based on factor analyses of the items included in the three scales investigated in the present study, we argue that a specific theoretical approach would be a better starting point. As an example of such approach we recently constructed a new workaholism scale, anchoring all its items within addiction theory before subjecting the items to psychometric testing (Andreassen *et al.*, 2012). Research should also be done to contrast and cross-validate workaholism measures with measures of related concepts such as job satisfaction (Staw and Cohen-Charash, 2005), job stress (Ganster and Schaubroeck, 1991), passion towards work (Vallerand *et al.*, 2010), work engagement (Shimazu and Schaufeli, 2009) and work over-commitment (Lehr *et al.*, 2010).

Conclusion

The present study showed that the long-term test-retest reliability of the workaholism instruments was good, suggesting that the construct assessed by these is quite stable over time. Future studies, however, should aim at investigating the degree to which workaholism is a stable personality trait, and the degree to which workaholism can be caused and changed by societal and organisational variables. The factor structure of the different workaholism measures has converged across studies to a small extent, and the present study did not support previously suggested factor structures, with the exception of the results from the explorative factor analyses of the DUWAS items. Future studies should therefore continue to investigate the factor structure of the workaholism instruments across different segments of workers, ages, genders and cultures.

References

Andreassen, C.S., Hetland, J. and Pallesen, S. (2010), "The relationship between workaholism, basic needs satisfaction at work and personality", *European Journal of Personality*, Vol. 24, pp. 3-17.

Andreassen, C.S., Ursin, H. and Eriksen, H.R. (2007), "The relationship between strong motivation to work, 'workaholism', and health", *Psychology & Health*, Vol. 22, pp. 615-629.

Andreassen, C.S., Griffiths, M.D., Hetland, J. and Pallesen, S. (2012), "Development of a work addiction scale", *Scandinavian Journal of Psychology*, Vol. 53, pp. 265-272.

Buelens, M. and Poelmans, S.A.Y. (2004), "Enriching the Spence and Robbins' typology of workaholism, demographic, motivational and organizational correlates", *Journal of Organizational Change Management*, Vol. 17, pp. 440-458.

Burke, R.J. (2001), "Workaholism in organizations: the role of organizational values", *Personnel Review*, Vol. 30, pp. 637-645.

Burke, R.J. (2006), "Workaholic types: it is not how hard you work but why you work hard", in Burke, R. (Ed.), *Research Companion to Working Time and Work Addiction*, Edward Elgar, Cheltenham, pp. 173-192.

Burke, R.J., Richardsen, A.M. and Martinussen, M. (2002), "Psychometric properties of Spence and Robbins' measures of workaholism components", *Psychological Reports*, Vol. 91, pp. 1098-1104.

Clark, L. (1996), *SNAP Manual for Administration, Scoring, and Interpretation*, University of Minnesota Press, Minneapolis, MN.

Costa, P.T. and McCrae, R.R. (1994), "Set like plaster? Evidence for the stability of the adult personality", in Heatheron, T.F. and Weinberger, J.L. (Eds), *Can Personality Change?*, American Psychological Associaton, Washintgton, DC, pp. 21-40.

del Libano, M., Llorens, S., Salanova, M. and Schaufeli, W. (2010), "Validity of a brief workaholism scale", *Psicothema*, Vol. 22, pp. 143-150.

Ersoy-Kart, M. (2005), "Reliability and validity of the workaholism battery (Work-BAT): Turkish form", *Social Behavior and Personality*, Vol. 33, pp. 609-618.

Flowers, C.P. and Robinson, B.E. (2002), "A structural and discriminant analysis of the work addiction risk test", *Educational and Psychological Measurement*, Vol. 62, pp. 517-526.

Ganster, D.C. and Schaubroeck, J. (1991), "Work stress and employee health", *Journal of Management*, Vol. 17, pp. 235-271.

Horn, J.L. (1965), "A rationale and a test for the number of factors in factor analysis", *Psychometrika*, Vol. 30, pp. 179-185.

Hu, L.T. and Bentler, P.M. (1999), "Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives", *Structural Equation Modeling*, Vol. 6, pp. 1-55.

Huang, J.C., Hu, C. and Wu, T.C. (2010), "Psychometric properties of the Chinese version of the workaholism battery", *Journal of Psychology*, Vol. 144, pp. 163-183.

Kanai, A., Wakabayashi, M. and Fling, S. (1996), "Workaholism among employees in Japanese corporations: an examination based on the Japanese version of the workaholism scales", *Japanese Psychological Research*, Vol. 38, pp. 192-203.

Lehr, D., Koch, S. and Hillert, A. (2010), "Where is (im)balance? Necessity and construction of evaluated cut-off points for effort-reward imbalance and overcommitment", *Journal of Occupational and Organizational Psychology*, Vol. 83, pp. 251-261.

Lonner, W.J. and Ibrahim, F.A. (1996), "Appraisal and assessment in cross-cultural counseling", in Pedersen, P.B. and Draguns, J.G. (Eds), *Counseling Across Cultures*, 4th ed., Sage Publications, Thousand Oaks, CA, pp. 292-322.

McMillan, L.H.W., O'Driscoll, M.P. and Burke, R.J. (2003), "Workaholism: a review of theory, research and new directions", in Cooper, C.L. and Robertson, I.T. (Eds), *International Review of Industrial and Organizational Psychology*, John Wiley, New York, NY, pp. 167-190.

McMillan, L.H.W., Brady, E.C., O'Driscoll, M.P. and Marsh, N.V. (2002), "A multifaceted validation study of Spence and Robbins' (1992) workaholism battery", *Journal of Occupational and Organizational Psychology*, Vol. 75, pp. 357-368.

McMillan, L.H.W., O'Driscoll, M.P., Marsh, N.V. and Brady, E.C. (2001), "Understanding workaholism: data synthesis, theoretical critique, and future design strategies", *International Journal of Stress Management*, Vol. 8, pp. 69-91.

Machlowitz, M. (1980), *Workaholics: Living with Them, Working with Them*, Addison-Wesley, Reading, MA.

Moes, G.S., Johnson, W.B. and Lall, R. (1996), "Personality characteristics of successful navy submarine personnel", *Military Medicine*, Vol. 161, pp. 239-242.

Mudrack, P.E. and Naugthon, T.J. (2001), "The assessment of workaholism as behavioral tendencies: scale development and preliminary empirical testing", *International Journal of Stress Management*, Vol. 8, pp. 93-111.

Ng, T.W.H., Sorensen, K.L. and Feldman, D.C. (2007), "Dimensions, antecedents, and consequences of workaholism: a conceptual integration and extension", *Journal of Organizational Behavior*, Vol. 28, pp. 111-136.

Oates, W. (1971), *Confessions of a Workaholic: The Facts About Work Addiction*, World, New York, NY.

Pedazur, E.J. and Schmelkin, L.P. (1991), *Measurement, Design and Analysis: An Integrated Approach*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y. and Podsakoff, N.P. (2003), "Common method biases in behavioral research: a critical review of the literature and recommended remedies", *Journal of Applied Psychology*, Vol. 88, pp. 879-903.

Porter, G. (1996), "Organizational impact of workaholism: suggestions for researching the negative outcomes of exessive work", *Journal of Occupational Health Psychology*, Vol. 1, pp. 70-84.

Robinson, B.E. (1989), *Workaholism: Hidden Legacies of Adult Children*, Health Communications, Deerfield Beach, FL.

Robinson, B.E. (1998), *Chained to the Desk: A Guidebook for Workaholics, their Partners and Children, and the Clinicians Who Treat Them*, New York University Press, New York, NY.

Robinson, B.E. (1999), "The work addiction risk test: development of a tentative measure of workaholism", *Perceptual and Motor Skills*, Vol. 88, pp. 199-210.

Robinson, B.E. (2001), "Workaholism and family functioning: a profile of familial relationships, psychological outcomes, and research considerations", *Conteporary Family Therapy*, Vol. 23, pp. 123-135.

Robinson, B.E. and Phillips, B. (1995), "Measuring workaholism. Content validity of the work addiction risk test", *Psychological Reports*, Vol. 77, pp. 657-658.

Robinson, B.E. and Post, P. (1994), "Validity of the work addiction risk test", *Perceptual and Motor Skills*, Vol. 78, pp. 337-338.

Robinson, B.E. and Post, P. (1995), "Split-half-reliability of the work addiction risk test: development of a measure of workaholism", *Psychological Reports*, Vol. 76, p. 1226.

Robinson, B.E., Flowers, C.P. and Carroll, J.J. (2001), "Work stress and marriage: a theoretical model examening the relationship between workaholism and marital cohesion", *International Journal of Stress Management*, Vol. 8, pp. 165-175.

Robinson, B.E., Post, P. and Khakee, J.F. (1992), "Test-retest reliability of the work addiction risk test", *Perceptual and Motor Skills*, Vol. 74, p. 926.

Schaufeli, W.B., Shimazu, A. and Taris, T.W. (2009), "Being driven to work excessively hard. The evaluation of a two-factor measure of workaholism in the Netherlands and Japan", *Cross-Cultural Research*, Vol. 43, pp. 320-348.

Schaufeli, W.B., van Wijhe, C., Peeters, M. and Taris, T. (2011), "Reek's psychological instruments. Work addiction, a concept measured", *Gedrag & Organisatie*, Vol. 24, pp. 43-63.

Scott, K.S., Moore, K.S. and Miceli, M.P. (1997), "An exploration of the meaning and consequences of workaholism", *Human Relations*, Vol. 50, pp. 287-314.

Shimazu, A. and Schaufeli, W.B. (2009), "Is workaholism good or bad for employee well-being? The distinctiveness of workaholism and work engagement among Japanese employees", *Industrial Health*, Vol. 47, pp. 495-502.

Snir, R. and Harpaz, I. (2009), "Cross-cultural differences concerning heavy work investment", *Cross-cultural Research*, Vol. 43, pp. 309-319.

Spence, J.T. and Robbins, A.S. (1992), "Workaholism – definition, measurement, and preliminary results", *Journal of Personality Assessment*, Vol. 58, pp. 160-178.

Staw, B.M. and Cohen-Charash, Y. (2005), "The dispositional approach to job satisfaction: more than a mirage, but not yet an oasis", *Journal of Organizational Behavior*, Vol. 26, pp. 59-78.

Streiner, D.L. and Norman, G.R. (2008), *Health Measurement Scales. A Pratical Guide to their Development and Use*, 4th ed., Oxford University Press, Oxford.

Taris, T.W., Schaufeli, W. and Shimazu, A. (2010), "The push and pull of work: the difference between workaholism and work engagement", in Bakker, A.B. and Leiter, M.P. (Eds), *Work Engagement. A Handbook of Essential Theory and Research*, Psychological Press, Hove, pp. 39-53.

Vallerand, R.J., Paquet, Y., Philippe, F.L. and Charest, J. (2010), "On the role of passion for work in burnout: a process model", *Journal of Personality*, Vol. 78, pp. 289-312.

Weir, J.P. (2005), "Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM", *Journal of Strength and Conditioning Research*, Vol. 19, pp. 231-240.

Woods, C.M., Oltmanns, T.F. and Turkheimer, E. (2008), "Detection of aberrant responding on a personality scale in a military sample: an application of evaluating person fit with two-level logistic regression", *Psychological Assessment*, Vol. 20, pp. 159-168.

Yen, M. and Lo, L-H. (2002), "Examining test-retest reliability: an intra-class correlation approach", *Nursing Research*, Vol. 51, pp. 59-62.

Corresponding author
Cecilie Schou Andreassen can be contacted at: cecilie.andreassen@psych.uib.no