

THE CONVERGENCE BETWEEN SELF AND OBSERVER RATINGS OF WORKAHOLISM: A COMPARISON BETWEEN COUPLES

ALESSANDRA FALCO
LUCA KRAVINA
DAMIANO GIRARDI
LAURA DAL CORSO
ANNAMARIA DI SIPIO
NICOLA A. DE CARLO
UNIVERSITY OF PADOVA

In the study of individual and behavioral characteristics, the need emerges to use observer ratings as well as self report type instruments in order to limit the intrinsic errors in both of the research methodologies. The present study looks deeper into the role of observer ratings in measuring workaholism. In particular, we evaluate the metric properties of a scale of observer rating of work addiction (DUWAS-OR) in terms of validity and reliability, and we determine its effectiveness in terms of distinguishing between workaholic workers. A questionnaire was administered to a sample of 243 couples of husband/wife or partners ($N = 486$) including scales of self report and observer rating of the partner, together with scales measuring workload and work-family conflict. A confirmatory factor analysis of the DUWAS-OR highlighted a two factor model of the scale, which presents good fit indexes; the analysis of the correlations between variables suggests good properties in terms of discriminant and convergent validity. Finally, using McNemar's test it was possible to determine that the number of participants identified as workaholic does not differ between the two methods, with a percentage of perfect correspondence of more than 80%. These results suggest the opportuneness of using an integrated approach to measure workaholism.

Key words: Workaholism; Observer ratings; Self rating; Integrated approach; Convergent and discriminant validity; Measure of workaholism.

Correspondence concerning this article should be addressed to Alessandra Falco, Dipartimento FISPPA – Sezione di Psicologia Applicata, Università di Padova, Via Venezia 8, 35131 PADOVA(PD), Italy. E-mail: alessandra.falco@unipd.it

INTRODUCTION

Traditionally, in the field of Industrial-Organizational Psychology (I-O Psychology), individual differences and behaviors are measured through self report instruments. This approach guarantees numerous advantages. In the first place, it is easy for the researcher or the practitioner to obtain a rating directly from the worker who is the object of the research (Oh, Wang, & Mount, 2011). Some individual dimensions are, moreover, so intimate that it is difficult for an external observer to evaluate them (e.g., emotional stability; Spain, Eaton, & Funder, 2000), which makes self report the most suitable instrument to reveal them. Finally, self report measures of individual differences are able to predict numerous job-related outcomes such as, for instance, job performances (Le et al., 2011), job satisfaction (Heller, Ferris, Brown, & Watson, 2009), job burnout

(Swider & Zimmerman, 2010) and psycho-physical strain (Löckenhoff, Duberstein, Friedman, & Costa, 2011).

However, some authors have criticized the exclusive use of self evaluation to measure individual differences (Oh et al., 2011). Colbert, Judge, Choi, and Wang (2012) identify two reasons why self report measures of individual differences can be biased: self deception, and faking. Judge, Erez, and Thoresen (2000) define self deception as an unconscious tendency to see oneself in a positive light, at the same time denying information that threatens the self. According to this perspective, individuals may not have the necessary self-insight to accurately evaluate their own traits. Faking refers to an intentional effort to respond in a socially desirable manner on a personality test with the aim of giving a positive image of oneself (Berry & Sackett, 2009). The tendency to distort (misrepresent) responses can be particularly evident in some subjects (e.g., subjects with low conscientiousness, high neuroticism; Morgeson et al., 2007), or in the case where self ratings occur in a competitive setting (e.g., personnel selection; Ployhart, 2006).

These considerations have led some authors to raise doubts about the validity of self report measures of individual differences (Morgeson et al., 2007; Oh et al., 2011).

A possible alternative suggested by the literature refers to the use of observer rating to reveal individual differences (Chang, Connelly, & Geeza, 2012; Colbert et al., 2012; Sackett & Lievens, 2008). While self report measures are influenced by the tendency of the respondents to present, consciously or unconsciously, a positive image of themselves, it is proposed that observer ratings are less or not at all affected by this bias. In the first place, observers should not be influenced by the self deception of the observed subject. Moreover, observers are not expected to be inclined to exaggerate the ratings of the individuals, as they would not obtain any benefit from inflating the test scores (Zimmerman, Triana, & Barrick, 2010).

The topic of observer rating is not new to I-O Psychology. It has already been used with good results, for example, to assess job stressors (e.g., job control, time pressure, job barriers; Grebner, Semmer, & Elfering, 2005; Greiner & Krause, 2006), psycho-physical strain (Falco et al., 2012; Waldenström et al., 2008), leadership (Fleenor, Smither, Atwater, Braddy, & Sturm, 2010), and job-related outcomes (e.g., job performance, Heidemeier & Moser, 2009; citizenship performance, Kamdar & Van Dyne, 2007).

Connelly and Ones (2010), in a meta-analysis on about 200 studies, examined the accuracy of observer ratings of individual differences, in terms of correlation between two or more observer ratings (i.e., interrater reliability), correlation between self and observer rating (i.e., convergence), and correlation between observer rating and relevant behavior and outcomes (job performance, i.e., criterion-related validity). In general, medium values emerged for interrater reliability and convergence (both between .39 and .51), once the effect of error measure had been controlled. These values vary in function of the difference observed (more or less easy to observe, i.e., visibility; more or less subject to social desirability, i.e., evaluativeness) and of the characteristics of the observer (level of familiarity of the observer with the observed subject, i.e., acquaintance).

These results are in line with those reported in another meta-analysis, carried out by Connolly, Kavanagh, and Viswesvaran (2007). This suggests that, although there is an overlap between constructs measured through self and observer rating (i.e., redundancy), there is, however, a quota of unique variance associated with each perspective, and that the differences between self ratings and observer ratings are substantive and not due to a statistical artifact (Zimmerman

et al., 2010). This can be explained in light of the fact that self and observer ratings capture different aspects of the individual. Self rating assesses the perception that an individual has of him/herself (i.e., individual's identity), while observer rating reveals the perception that others have of the individual, which is primarily based on the individual's previous behavior (i.e., reputation; Tett & Burnett, 2003).

Finally, Connelly and Ones (2010) examined the relation between observer rating of individual differences and job performance, evaluated by an independent observer (different from the one who carried out the evaluations of personality/individual differences). The results were compared with those emerging in previous meta-analyses, in which individual differences were assessed using self report (Barrick, Mount, & Judge, 2001). More intense correlations emerged between job performance and observer rating of individual differences with respect to those between job performance and self rating. The authors concluded that observer rating provides a stronger validity for predicting job performance than self report measures, and that the studies carried out up to now, based prevalently on self-report of individual differences, have underestimated the relation between individual differences and job related outcomes.

Furthermore, there is a rising interest for observer rating, especially for diagnosis, in the evaluation of behavioral addictions, such as — for example — pathological gambling, compulsive buying, and compulsive exercise (Albrecht, Kirschner, & Grüsser, 2007; Grant, Levine, Kim, & Potenza, 2005; van Elburg, Hoek, Kas, & van Engeland, 2007).

However, also observer rating is not devoid of criticism (Colbert et al., 2012). The external expression of an individual difference may be influenced by external factors such as, for example, the cultural norms existing in a particular social context. For example, in evaluating work addiction it is possible that supervisors or the whole organization evaluate work addicted type behaviors positively, in that they are considered effective in reaching organizational aims (Ng, Sorensen & Feldman, 2007). This can therefore encourage the expression of these behaviors.

Moreover, situational cues are needed to trigger the (observable) expression of a latent individual difference, without which it is difficult for the evaluated trait to express itself and as a consequence be evaluated. It is also possible that observer rating, like self report, is subject to systematic distortions (Oh et al., 2011). For example, it may happen that observers intentionally minimize socially undesirable characteristics of the worker (e.g., being anxious) or exaggerate socially desirable characteristics (e.g., being a hardworker). Finally, the observer might attribute the behavior of the observed individual to dispositional more than situational factors (i.e., fundamental attribution error; Ross, 1977). This can result in an overestimation of desirable traits when the observed individual achieves good results (e.g., high job performance) or, vice versa, to an underestimation of desirable traits when the observed individual achieves poor results, even if these results (good and negative) can be attributed to situational factors, and therefore out of the individual's control. Within the I-O Psychology a construct that could have benefit from the combination between self-report and observer ratings is workaholism, defined by Schaufeli, Taris, and Bakker (2008) as “the tendency to work excessively hard in a compulsive way” (p. 204). Even though the literature on the topic is now particularly vast (e.g., Ng et al., 2007; Schaufeli, Bakker, van der Heijden, & Prins, 2009; Schaufeli, Taris, & Bakker, 2006) the usually does not include studies using the combination of self report and observer ratings of the construct, nor studies which evaluate workaholism using different self report instruments, with two significant exceptions. The first study, by Robinson and Kelley (1998), uses a modified version

of the Work Addiction Risk Test-WART (Robinson, 1989) so that participants can retrospectively evaluate the level of workaholism of their own parents: a limit of the study, recognized by the authors themselves, is that the WART was not designed for observer rating of workaholism. The second study (Chamberlin & Zang, 2009), uses the only scale of observer rating of workaholism present in the literature, the CWST — The Children of Workaholic Parents Test (Robinson & Carroll, 1999), whose psychometric characteristics, however, cannot be traced or evaluated. A limit of Chamberlin and Zang's study is that the scales, both of self report and observer rating, are compiled by the participants, who are requested to judge themselves and their parents. So, despite the attempt at combining methods, a single source of rating was used.

In this context, the present study proposes to look deeper into the role of observer rating in measuring workaholism, through the involvement of a sample of couples (husband/wife or partners). The choice of a measure of workaholism within the family and not inside an organization (e.g., considering the observer rating by colleagues) is due to the fact that the excesses of workaholism are likely to be observed to a greater extent outside the workplace. On the one hand, in fact, Ng et al. (2007) underline that often workaholic behaviors are not discouraged by organizations, on the other it is easy to imagine that it is precisely the husband/wife or partner who is in a privileged position to witness the typical behavior of a workaholic, such as, for instance, working weekends, always coming back home very late, not being able to relax or dedicate time to activities outside of work.

OBJECTIVES

The aim of the present work consists of testing the psychometric properties of the observer rating version (DUWAS-OR) of the DUWAS-R scale (Kravina, Falco, Girardi, & De Carlo, 2010). The DUWAS-R scale represents the revised Italian version of the DUWAS – *Dutch Workaholism Scale* (Schaufeli et al., 2006), which keeps stable the two factor structure of the original scale (composed of two subscales: Work Excessively – WE, and Work Compulsively – WC), but differentiates itself in terms of the number of items and the type of response scale used. In particular, we expect: a) the structure of the observer rating scale of workaholism (DUWAS-OR) to mirror the two factor structure of the self report scale (DUWAS-R), described in Kravina et al. (2010); b) the DUWAS-OR scale to provide good results in terms of reliability and validity; c) that the number of participants identified as workaholic do not vary when the method used to assess them varies (self report vs. observer rating).

METHOD

Participants and Procedure

A questionnaire was administered to a convenience sample of 243 couples (husband and wife/partners), corresponding to a total of $N = 486$ participants, including the scales described in the section Materials. Personal information concerning participants, including gender, age, qualifications and work position, are summarized in Table 1.

TABLE 1
Personal details of the participants in the study ($N = 486$)

Variable	Frequency	% valid	Missing values
<i>Gender</i>	486	–	0
Female	243	50.0	–
Male	243	50.0	–
<i>Educational level</i>	482	–	4
Compulsory school certificate	132	27.4	–
High school diploma	215	44.6	–
University degree	135	28.0	–
<i>Position held</i>	454	–	32
Top-level manager	77	17.0	–
Mid-level manager	47	10.3	–
Office worker	157	34.6	–
Blue-collar	102	22.5	–
Other kinds of employment (e.g., freelancer)	71	15.6	–
	<i>N</i>	<i>Mean</i>	<i>SD</i>
<i>Age</i>	486	54.7	6.72

Every member of the couple was asked to compile the self report scales of workaholism, workload, and work-family conflict, so as the scale of observer rating of work addiction their partner. Regarding this last scale, explicit instructions were given to, each member of the couple of spouses/partners asking them to evaluate their own husband or wife/partner. A necessary condition to take part in the sample was to be currently employed in paid work: therefore the couples in which one or both of the people were pensioners, housewives or unemployed were excluded a priori from the sample.

The questionnaire was compiled anonymously in the presence of a researcher. Each couple was assigned a specific code number, in a way that for each participant the data of the self-evaluation and those of the hetero-evaluation given by their partner would be on the same row of the data-set. The data deriving from self report and the data from the evaluation compiled by the husband/wife or partner correspond to the same line of the data-set.

Materials

The questionnaire administered to the participants was made up of four sections: the first was dedicated to the self report of workaholism; the second included the observer rating of workaholism by the husband or wife/partner; the third was aimed at detecting some possible effects of work addiction, such as workload and work-family conflict at an individual level; the fourth, finally, was dedicated to collecting some personal details of the participants.

Workaholism — self report. Self report of workaholism was performed using the DUWAS-R scale (Kravina, Falco, Girardi, & De Carlo, 2010). The scale includes 10 items, six concerning Work Excessively (Self-WE) and four Work Compulsively (Self-WC), using a 6-

point Likert scale of agreement/disagreement (1 = *strongly disagree*; 6 = *strongly agree*). Cronbach's alpha for the two dimensions was respectively .76 e .81.

Workaholism — observer rating. Observer rating was performed using the DUWAS-OR scale, also made up of 10 items, six regarding Work Excessively (Observer-WE) and four Work Compulsively (Observer-WC), using a 6-point Likert scale of agreement/disagreement (1 = *strongly disagree*; 6 = *strongly agree*). The items of the DUWAS-OR scale were obtained by reformulating the original items of the DUWAS-R scale in such a way as to allow the observer to assign a score, with respect to the proposed scale, considering his/her own husband or wife/partner. The items, reported in Table 2 subdivided by dimension, were preceded by specific instructions in which participants were asked to make reference to their own husband or wife/partner.

Workload. Workload was measured through 15 items (e.g., “Your workload requires you to work very quickly”) on a 6-point Likert scale (1 = *strongly disagree*; 6 = *strongly agree*). Cronbach's alpha was equal to .89.

Work-Family Conflict (WFC). Work-family conflict was assessed using two items (e.g., “I dedicate too little time to my family because of my work”) on a 6-point Likert scale (1 = *strongly disagree*; 6 = *strongly agree*). Cronbach's alpha was .73.

Demographic data. All participants were asked to indicate, in a specific section of the questionnaire, personal details such as gender, age, qualifications and position held.

TABLE 2
Items for observer rating ordered by dimension

Item	Dimension: Work Excessively (WE)
Item-WE1_OR	He/she is always in a hurry and seems to be racing against the clock
Item-WE2_OR	He/she dedicates much more time to work than to friends and hobbies
Item-WE3_OR	He/she is excessively involved in his/her work, beyond his/her possibilities
Item-WE4_OR	When working he/she imposes personal deadlines so as to keep him/herself under pressure
Item-WE5_OR	He/she simultaneously does many things, for example answering the telephone and making appointments during the lunch break
Item-WE6_OR	He/she continues working even after his/her colleagues have already left
Item	Dimension: Work Compulsively (WC)
Item-WC1_OR	Working with commitment is for him/her is an obligation, even in cases when he/she does not like what he/she is doing
Item-WC2_OR	It seems that he/she has an internal drive to work hard: a sensation that it is something that he/she has to do whether he/she wants to or not
Item-WC3_OR	There is something inside him/her that pushes him/her to work hard
Item-WC4_OR	He/she cannot refrain from always working with great commitment even when this is not pleasurable

Statistical Analyses

The psychometric properties of the DUWAS-OR scale were evaluated through a series of statistical analyses. Firstly, confirmatory factor analysis was carried out using the Lisrel 8.8 soft-

ware (Jöreskog & Sörbom, 1993). We assessed model fit using the χ^2 test, the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), the Non-Normed Fit Index (NNFI), and the Standardized Root Mean Square Residual (SRMR). Since the χ^2 is affected by sample size, the use of additional fit indexes is recommended. Values close to or smaller than .08 for RMSEA/SRMR and values close to or greater than .95 for CFI/NNFI indicate an acceptable model fit (Schermerle-Engel, Moosbrugger, & Müller, 2003). The internal consistency reliability of the DUWAS-OR was assessed by Cronbach's alpha.

Then, for each of the two dimensions into which workaholism is articulated, the coefficient Average Variance Extracted (AVE) was calculated, which represents the average amount of variation that a latent construct is able to explain in the observed variables to which it is theoretically related.

AVE can be used to assess both convergent and discriminant validity. A good convergent validity is verified when all indicators load significantly on their respective latent construct. AVE scores equal to or higher than .50 for each dimension indicate a good convergent validity (Bagozzi & Yi, 1988). In fact, if AVEs were lower than .50, the variance due to measurement error would be higher than the variance captured by the construct (Fornell & Larcker, 1981). In addition, the two dimensions can be considered distinct (i.e., discriminant validity) if the AVE of each of them is higher than the squared correlation between the two dimensions (shared variance, Fornell & Larcker, 1981).

In order to assess the criterion-related validity of the DUWAS-OR, the relations between the two dimensions of workaholism and some constructs were examined, such as workload and work-family conflict, indicated in the literature as possible consequences of workaholism. In particular, workaholic workers, who take on a considerable amount of work driven by the need to be constantly busy in some work activity, result as being particularly inclined to reporting perceived high levels of workload (Kanai & Wakabayashi, 2001; Kanai, Wakabayashi, & Fling, 1996; Schaufeli et al., 2009). With reference to work-family conflict, as a result of the great amount of time dedicated to working activities and the incapacity to maintain a balance between private life and working life, workaholic workers report high levels of WFC (Andreassen, Hetland, & Pallesen, 2012; Bonebright, Clay, & Ankenmann, 2000; Taris, Schaufeli, & Verhoeven, 2005).

Finally, to assess the correspondence between self report and observer rating, the McNemar's test was performed, which is useful in establishing whether two dichotomous categorical variables measure the same characteristic. In this specific case, the test was performed to check whether the number of participants classified as workaholic and non workaholic remains unchanged when the two methodologies, self report and observer rating, were used.

Results

With the aim of testing the two factor structure of the DUWAS-OR scale, a confirmatory factor analysis was carried out. For the DUWAS-OR scale (observer rating) we decided to re-test the same model which emerged for the DUWAS-R scale (Kravina et al., 2010) in which the covariance between error terms relative to item 8 and item 9 was freely estimated. This is justifiable from a theoretical point of view, since the two items have similar wording (see Table 2). Since the content of the items was not modified from the DUWAS-R to the DUWAS-OR scale, but on-

ly their formulation, so to make them suitable for observer rating, the two instruments can be considered as having the same factor structure.

On the whole, the fit indexes show an acceptable fit of the hypothesized theoretical model to the data: $\chi^2(33, N = 486) = 164.52, p < .001$; RMSEA = .089; CFI = .97; NNFI = .96; SRMR = .044. A value slightly above the threshold can be seen for RMSEA. The other indexes, however, converge in indicating an acceptable fit. The standardized factor loadings ranged from .61 to .86. The correlation between the two latent variables of the DUWAS-OR scale was high ($\Phi_{21} = .69$) and substantially in line with the correlation regarding the DUWAS-R scale ($\Phi_{21} = .71$) (Figure 1).

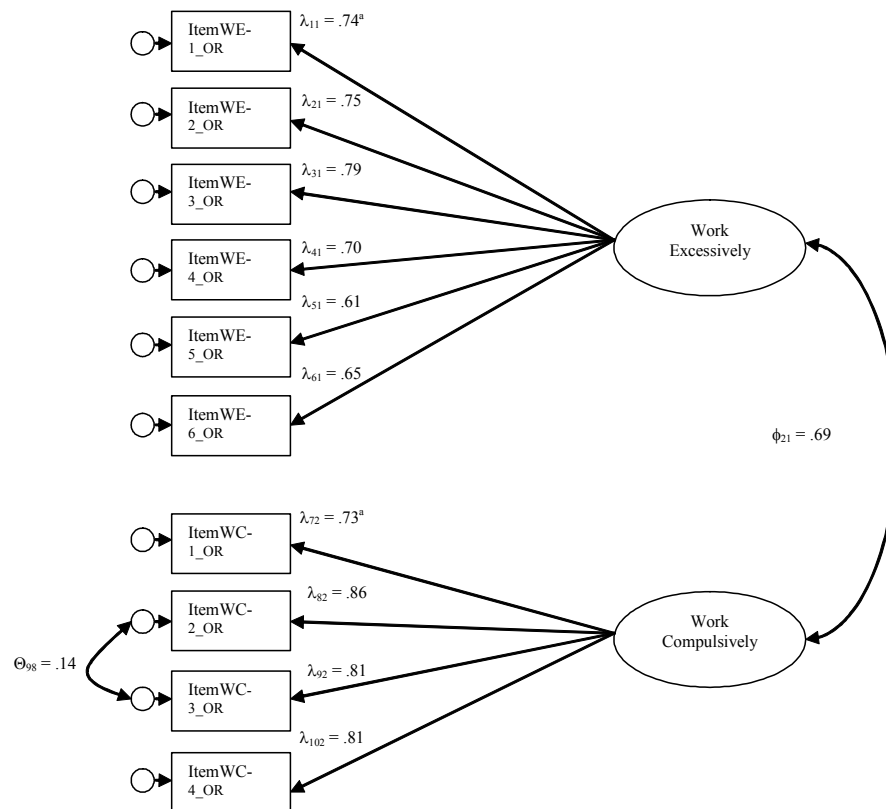


FIGURE 1
Model for the observer rating scale, with 10 items and two latent variables.

It is important to establish whether the two dimensions, from the perspective of the observer, are distinct and do not overlap. Therefore, the fit of the two factor model was compared to that of a single factor model using the χ^2 difference test (χ^2_D). The single factor model, more parsimonious, was specified fixing at 1 the correlation between the two latent factors (Brown, 2006). This correlation was instead estimated freely in the two factor model, which is less parsimonious. The two factor model presents a better fit than the single factor model, $\chi^2_D(1, N = 486) = 269.74, p < .001$, and is, therefore, preferable to the latter one.

The reliability of the two sub-scales, assessed using Cronbach's alpha, was satisfactory. The alpha coefficient was .86 for the WE dimension, and .89 for the WC dimension.

Average variance extracted was .50 for WE and .65 for WC. The two latent dimensions explained on average at least 50% of the variance in their measurement items (i.e., 50.3% for WE, 64.6% for WC). The AVE for each subscale was, moreover, higher than the shared variance between the two latent factors (.50 vs. .48 for WE; .65 vs. .48 for WC). Overall, the DUWAS-OR scale shows satisfactory convergent and discriminant validity.

As shown by the correlation matrix (Table 3), a medium intensity convergence between self reports and observer ratings of workaholism ($r = .49, p < .001$ for WE; $r = .43, p < .001$ for WC) was found. These values are in line with those emerging from the already cited study by Burke and Ng (2007) and in previous studies concerning individual differences (Connolly et al., 2007; Muck, Hell, & Gosling, 2007).

TABLE 3
Means, standard deviations and correlations between variables

	<i>M</i>	<i>SD</i>	1.	2.	3.	4.	5.	6.
1. Self-WE	3.92	1.04	—					
2. Self-WC	4.15	1.11	.50***	—				
3. Observer-WE	3.75	1.17	.49***	.29***	—			
4. Observer-WC	4.23	1.14	.30***	.43***	.60***	—		
5. Workload	4.07	0.93	.34***	.18**	.24***	.20***	—	
6. WFC	3.23	1.44	.32***	.13*	.28***	.15**	.38***	—

* $p < .05$. ** $p < .01$. *** $p < .001$.

Moreover, similar correlational patterns are found between workaholism and job stressors when the method used to reveal them is varied (self report vs. observer rating). As can be expected, the WFC correlated positively with workaholism. In particular, the WFC correlated more intensely with WE than with WC. This is true both in the case of self report ($r = .32, p < .001$ for Self-WE; $r = .13, p = .012$ for Self-WC) and observer rating ($r = .28, p < .001$ for Observer-WE; $r = .15, p = .005$ for Observer-WC). An analogous pattern, although less evident, emerges in relation to workload, which correlated more intensely with WE than with WC. In this case the difference is more marked in the case of self report ($r = .34, p < .001$ for Self-WE; $r = .18, p = .001$ for Self-WC) than for observer rating ($r = .24, p < .001$ for Observer-WE; $r = .20, p < .001$ for Observer-WC). Positive correlations between the dimensions of workaholism, workload and WFC can be found in other studies that measure workaholism as it is considered in the DUWAS scale (Schaufeli et al., 2009).

Identification of Workaholics

As pointed out by Schaufeli et al. (2009), in fact, to be able to talk about workaholism, a combination of high values is necessary in both of the dimensions. For this reason the workaholic participants were identified using the 75th percentile as the cut off criterion. The same cut off cri-

terion was also used for the observer rating scale. Concerning the self report scale (Work Excessively: $M = 3.92$, $SD = 1.04$; Work Compulsively: $M = 4.15$, $SD = 1.11$) the cut off values were, for the two dimensions respectively, 4.67 and 5.00. For the observer rating scale (Work Excessively: $M = 3.75$, $SD = 1.04$; Work Compulsively: $M = 4.23$, $SD = 1.11$), the cut off values were identical to those of self report (4.67 and 5.00 for the two dimensions, respectively).

The percentage of workers identified as workaholic, reported in Table 4, varied between 17.1% (self report) and 20% (observer rating). These values are in line with what is reported in the literature regarding the percentage of workaholics identified in the different samples even when using different measure instruments (Sussman, Lisha, & Griffiths, 2011).

TABLE 4
Workaholic identified through self report and observer rating

	<i>n</i>	%
<i>Self report</i>		
Non workaholic	403	82.9
Workaholic	83	17.1
<i>Observer rating</i>		
Non workaholic	389	80.0
Workaholic	97	20.0

Through McNemar's test (Pallant, 2010) it was possible to analyze the correspondence between self report and observer rating (Table 5). The associated probability was $p = .165$. The non significance of McNemar's test allows us to affirm that the percentage of participants determined as workaholic (17.1% with self report; 20.0% through observer rating) does not vary when the method used varies.

TABLE 5
Correspondence between self report and observer rating

	Non workaholic observer rating	Workaholic observer rating	Total
Non workaholic self report	352	51	403
Workaholic self report	37	46	83
Total	389	97	486

Moreover, through a cross-tabulation between the participants identified as non workaholic ($n = 352$) and workaholic ($n = 46$) by both of the instruments, it is possible to observe that 81.9% of them was identified in the same way through self report and observer rating. The participants in which there was no correspondence between self report and observer rating were there-

fore the 18.1%, which represents the percentage of participants wrongly classified in the methods combination. Of this number, 57.9% were participants labeled as workaholic by their own husband or wife/partner and not by themselves, and 42.1% were participants who were labeled as non workaholic by their own husband or wife/partner but defined themselves, through self report, as workaholic.

DISCUSSION

The first two aims of the study were to test the metric properties of an observer rating scale of workaholism. This scale, named DUWAS-OR, consists of an adaptation of the self report scale DUWAS-R, whose metric properties have been tested in a previous work (Kravina et al., 2010). The integrated use of self and observer rating of workaholism is in line with what is suggested in the literature, according to which this procedure allows the different angles of the phenomenon to be captured (Colbert et al., 2012). The evaluation carried out by a single rater, whether the participant in a study is called upon to assess him/herself or a third person (e.g., the supervisor, a colleague, partner), is affected by the individual perspective of the assessor/rater. This leads to a systematic distortion in measuring the dimension being examined. The integrated approach therefore helps to grasp different aspects of the individual and to reduce the effect of common method bias, which can lead to overestimating the correlations between the constructs (Chang et al., 2012; Conway & Lance, 2010).

Overall, the DUWAS-OR scale shows good psychometric properties. Confirmatory factor analysis allowed us to confirm the two factor structure, which mirrors that of the original DUWAS-R. In addition, DUWAS-OR evidenced good properties in terms of reliability, as well as convergent and discriminant validity. The two dimensions of workaholism, revealed through observer rating are, moreover, correlated with WFC and workload. These results replicate substantially those obtained through self report and are in line with what had previously emerged in the literature (e.g., Bonebright et al., 2000; Schaufeli et al., 2009).

The third aim of the study consists in verifying whether the number of workers identified as workaholic is different when the method of evaluation varies.

Statistical analyses revealed how the number of workaholics does not differ in a statistically significant manner with the variation of the used method (self-report and observer rating): in particular, it is possible to find a perfect correspondence in more than 80% of the cases within workers that can be defined as workaholic or non workaholic through the instrument of self-report and labeled the same way by their partner.

Thus, in general, it is possible to affirm that the DUWAS-OR scale is an effective instrument for observer rating of workaholism, given both the goodness of the psychometric characteristics in terms of factor structure, discriminant validity and convergent validity and in terms of correspondence between self report and observer rating.

It is possible to notice some limits in the present study. First of all, the sample used is a convenience sample which does not allow us to generalize results, meaning that further studies are required to evaluate the goodness of the observer rating instrument. Secondly, the questionnaire administered does not contemplate, among the effects and the consequences of workaholism, burnout, which is considered one of the main consequences of work addiction

(Schaufeli et al., 2009). For providing more proofs about the criterion-related validity of the DUWAS-OR scale, a new study could therefore examine the relation between workaholism and burnout.

Thirdly, the participants come from different organizational realities and despite the simultaneous use of methodologies of self report and observer rating, it was not possible to obtain objective indicators, such as, for example, the quantity of hours worked, absences and days off; this prevents us from completing a process of triangulation of workaholism. New studies which simultaneously evaluate workaholism using different instruments will be able to improve knowledge about the examined construct and to provide new evidence about the goodness of the instrument presented in this study.

Fourthly, as pointed out by Chang et al. (2012), in order to limit the possible biases deriving from the observer's evaluation, it is advisable to consider the evaluation of at least two external observers. Further research involving more than one observer simultaneously, both within the family and from the working environment will, therefore, be able to further improve the integrated approach in measuring workaholism.

REFERENCES

- Albrecht, U., Kirschner, N. E., & Grüsser, S. M. (2007). Diagnostic instruments for behavioural addiction: An overview. *GMS Psycho-Social-Medicine*, 4, 1-11.
- Andreassen, C. S., Hetland, J., & Pallesen, S. (2012). Workaholism and work-family spillover in a cross-occupational sample. *European Journal of Work and Organizational Psychology*. Advance online publication. doi:10.1080/1359432X.2011.626201
- Bagozzi, R., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16(1), 74-94. doi:10.1007/BF02723327
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9(1-2), 9-30. doi:10.1111/1468-2389.00160
- Berry, C. M., & Sackett, P. R. (2009). Faking in personnel selection: Tradeoffs in performance versus fairness resulting from two cut-score strategies. *Personnel Psychology*, 62(4), 833-863. doi:10.1111/j.1744-6570.2009.01159.x
- Bonebright, C. A., Clay, D. L., & Ankenmann, R. D. (2000). The relationship of workaholism with work-life conflict, life satisfaction, and purpose in life. *Journal of Counseling Psychology*, 47(4), 469-477. doi:10.1037/0022-0167.47.4.469
- Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Burke, R. J., & Ng, E. S. W. (2007). Workaholic behaviors: Do colleagues agree? *International Journal of Stress Management*, 14(3), 312-320. doi:10.1037/1072-5245.14.3.312
- Chamberlin, C. M., & Zhang, N. (2009). Workaholism, health, and self-acceptance. *Journal of Counseling & Development*, 87(2), 159-169. doi:10.1002/j.1556-6678.2009.tb00563.x
- Chang, L., Connelly, B. S., & Geeza, A. A. (2012). Separating method factors and higher order traits of the Big Five: A meta-analytic multitrait-multimethod approach. *Journal of Personality and Social Psychology*, 102(2), 408-426. doi:10.1037/a0025559
- Colbert, A. E., Judge, T. A., Choi, D., & Wang, G. (2012). Assessing the trait theory of leadership using self and observer ratings of personality: The mediating role of contributions to group success. *The Leadership Quarterly*, 23(4), 670-685. doi:10.1016/j.leaqua.2012.03.004
- Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136(6), 1092-1122. doi:10.1037/a0021212
- Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment*, 15(1), 110-117. doi:10.1111/j.1468-2389.2007.00371.x
- Conway, J., & Lance, C. (2010). What reviewers should expect from authors regarding common method bias in organizational research. *Journal of Business and Psychology*, 25(3), 325-334. doi:10.1007/s10869-010-9181-6
- Falco, A., Girardi, D., Sarto, F., Marcuzzo, G., Vianello, L., Magosso, D., et al. (2012). Una nuova scala di misura degli effetti psico-fisici dello stress lavoro-correlato in una prospettiva d'integrazione di me-

- todi [A new scale for measuring the psycho-physical effects of work-related stress in a perspective of methods integration]. *La Medicina del Lavoro*, 103(4), 276-287.
- Fleenor, J. W., Smither, J. W., Atwater, L. E., Braddy, P. W., & Sturm, R. E. (2010). Self-other rating agreement in leadership: A review. *The Leadership Quarterly*, 21(6), 1005-1034. doi:10.1016/j.leaqua.2010.10.006
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50. doi:10.2307/3151312
- Grant, J. E., Levine, L., Kim, D., & Potenza, M. N. (2005). Impulse control disorders in adult psychiatric inpatients. *American Journal of Psychiatry*, 162, 2184-2188.
- Grebner, S., Semmer, N. K., & Elfering, A. (2005). Working conditions and three types of well-being: A longitudinal study with self-report and rating data. *Journal of Occupational Health Psychology*, 10(1), 31-43. doi:10.1037/1076-8998.10.1.31
- Greiner, B. A., & Krause, N. (2006). Observational stress factors and musculoskeletal disorders in urban transit operators. *Journal of Occupational Health Psychology*, 11(1), 38-51. doi:10.1037/1076-8998.11.1.38
- Heidemeier, H., & Moser, K. (2009). Self-other agreement in job performance ratings: A meta-analytic test of a process model. *Journal of Applied Psychology*, 94(2), 353-370. doi:10.1037/0021-9010.94.2.353
- Heller, D., Ferris, D. L., Brown, D., & Watson, D. (2009). The influence of work personality on job satisfaction: Incremental validity and mediation effects. *Journal of Personality*, 77(4), 1051-1084. doi:10.1111/j.1467-6494.2009.00574.x
- Jöreskog, K. G., & Sörbom, D. (1993). *Lisrel 8: Structural equation modeling with the SIMPLIS command language*. Hillsdale, N.J.: Lawrence Erlbaum Publishers.
- Judge, T. A., Erez, A., & Thoresen, C. J. (2000). Why negative affectivity (and self-deception) should be included in job stress research: Bathing the baby with the bath water. *Journal of Organizational Behavior*, 21(1), 101-111. doi:10.1002/(SICI)1099-1379(200002)21:1<101::AID-JOB966>3.0.CO;2-Q
- Kamdar, D., & Van Dyne, L. (2007). The joint effects of personality and workplace social exchange relationships in predicting task performance and citizenship performance. *Journal of Applied Psychology*, 92(5), 1286-1298. doi:10.1037/0021-9010.92.5.1286
- Kanai, A., & Wakabayashi, M. (2001). Workaholism among Japanese blue-collar employees. *International Journal of Stress Management*, 8(2), 129-145. doi:10.1023/A:1009529314121
- Kanai, A., Wakabayashi, M., & Fling, S. (1996). Workaholism among employees in Japanese corporations: An examination based on the Japanese version of the Workaholism Scales. *Japanese Psychological Research*, 38(4), 192-203. doi:10.1111/j.1468-5884.1996.tb00024.x
- Kravina, L., Falco, A., Girardi, D., & De Carlo, N. A. (2010). Workaholism among management and workers in an Italian cooperative enterprise. *TPM – Testing, Psychometrics, Methodology in Applied Psychology*, 17(4), 201-216.
- Le, H., Oh, I.-S., Robbins, S. B., Ilies, R., Holland, E., & Westrick, P. (2011). Too much of a good thing: Curvilinear relationships between personality traits and job performance. *Journal of Applied Psychology*, 96(1), 113-133. doi:10.1037/a0021016
- Löckenhoff, C. E., Duberstein, P. R., Friedman, B., & Costa, P. T. (2011). Five-factor personality traits and subjective health among caregivers: The role of caregiver strain and self-efficacy. *Psychology and aging*, 26(3), 592-604. doi:10.1037/a0022209
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60(3), 683-729. doi:10.1111/j.1744-6570.2007.00089.x
- Muck, P. M., Hell, B., & Gosling, S. D. (2007). Construct validation of a short five-factor model instrument: A self-peer study on the German adaptation of the Ten-Item Personality Inventory (TIPI-G). *European Journal of Psychological Assessment*, 23(3), 166-175. doi:10.1027/1015-5759.23.3.166
- Ng, T. W. H., Sorensen, K. L., & Feldman, D. C. (2007). Dimensions, antecedents, and consequences of workaholism: A conceptual integration and extension. *Journal of Organizational Behavior*, 28, 111-136. doi:10.1002/job.424
- Oh, I.-S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology*, 96(4), 762-773. doi:10.1037/a0021832
- Pallant, J. (2010). *SPSS Survival Manual*. New York, NY: McGraw-Hill
- Ployhart, R. E. (2006). Staffing in the 21st century: New challenges and strategic opportunities. *Journal of Management*, 32(6), 868-897. doi:10.1177/0149206306293625
- Robinson, B. E. (1989). *Work Addiction: Hidden legacies of adult children*. Deerfield Beach, FL: Health Communications.
- Robinson, B. E., & Carroll, J. J. (1999). Assessing the offspring of workaholic parents: The children of workaholics screening test. *Perceptual and Motor Skills*, 88(3c), 1127-1134. doi:10.2466/pms.1999.88.3c.1127
- Robinson, B. E., & Kelley, L. (1998). Adult children of workaholics: Self-concept, anxiety, depression, and locus of control. *The American Journal of Family Therapy*, 26(3), 223-238. doi:10.1080/01926189808251102

- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology*, 10, 173-220. doi:10.1016/S0065-2601(08)60357-3
- Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology*, 59, 419-450. doi:10.1146/annurev.psych.59.103006.093716
- Schaufeli, W. B., Bakker, A. B., Van Der Heijden, F. M. M. A., & Prins, J. T. (2009). Workaholism among medical residents: It is the combination of working excessively and compulsively that counts. *International Journal of Stress Management*, 16(4), 249-272. doi:10.1037/a0017537
- Schaufeli, W. B., Taris, T. W., & Bakker, A. (2006). Dr. Jekyll and Mr. Hide: On the differences between work engagement and workaholism. In R. Burke (Ed.), *Research companion to working time and work addiction* (pp. 193-217). Northampton, MA: Edward Elgar.
- Schaufeli, W. B., Taris, T. W., & Bakker, A. (2008). It takes two tango: Workaholism is working excessively and working compulsively. In R. J. Burke & C. L. Cooper (Eds.), *The long work hours culture: causes, consequences, and choices* (pp. 203-225). Bingley, UK: Emerald.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, 8(2), 23-74. SIMPLIS command language. Hillsdale, NJ: Lawrence Erlbaum Publishers.
- Spain, J. S., Eaton, L. G., & Funder, D. C. (2000). Perspectives on personality: The relative accuracy of self versus others for the prediction of emotion and behavior. *Journal of Personality*, 68(5), 837-867. doi:10.1111/1467-6494.00118
- Sussman, S., Lisha, N., & Griffiths, M. (2011). Prevalence of the addictions: A problem of the majority or the minority? *Evaluation & the Health Professions*, 34(1), 3-56. doi:10.1177/0163278710380124
- Swider, B. W., & Zimmerman, R. D. (2010). Born to burnout: A meta-analytic path model of personality, job burnout, and work outcomes. *Journal of Vocational Behavior*, 76(3), 487-506. doi:10.1016/j.jvb.2010.01.003
- Taris, T. W., Schaufeli, W. B., & Verhoeven, L. C. (2005). Workaholism in the Netherlands: Measurement and implications for job strain and work-nonwork conflict. *Applied Psychology*, 54(1), 37-60. doi:10.1111/j.1464-0597.2005.00195.x
- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, 88(3), 500-517. doi:10.1037/0021-9010.88.3.500
- van Elburg, A. A., Hoek, H. W., Kas, M. J. H., & van Engeland, H. (2007). Nurse evaluation of hyperactivity in anorexia nervosa: A comparative study. *European Eating Disorders Review*, 15(6), 425-429. doi: 10.1002/erv.803
- Waldenström, K., Ahlberg, G., Bergman, P., Forsell, Y., Stoetzer, U., Waldenström, M., & Lundberg, I. (2008). Externally assessed psychosocial work characteristics and diagnoses of anxiety and depression. *Occupational and Environmental Medicine*, 65(2), 90-96. doi:10.1136/oem.2006.031252
- Zimmerman, R. D., Triana, M. C., & Barrick, M. R. (2010). Predictive criterion-related validity of observer ratings of personality and job-related competencies using multiple raters and multiple performance criteria. *Human Performance*, 23(4), 361-378. doi:10.1080/08959285.2010.50104