**Hofmans, J., Dries, N., Pepermans, R. (2008). The career satisfaction scale:
Response bias among men and women. Journal of Vocational Behavior, 73 (3),
397-403.**

Joeri HOFMANS (corresponding author)
Katholieke Universiteit Leuven
Department of Psychology
Tiensestraat 102
3000 Leuven
Belgium
Tel: +32.16.32.61.08.
Fax: +32.16.32.59.93.
Email: joeri.hofmans@psy.kuleuven.be

Nicky DRIES & Roland PEPERMANS
Vrije Universiteit Brussel
Work, Organizational and Economic Psychology
Pleinlaan 2
1050 Brussel
Belgium
Tel: +32.2.629.39.63.
Fax: +32.2.629.39.48.
E-mail: nicky.dries@vub.ac.be; roland.pepermans@vub.ac.be

**Response behavior and the career satisfaction scale: Testing for gender
invariance**

Gender is a specific predictor of interest in many studies on career success. Important implications, mostly about “glass ceilings” and discriminatory career opportunity structures within organizations on a micro-level and societies on a macro-level, are often drawn from these studies. However, when comparing subjective career success (i.e. career satisfaction, which is a latent variable) between men and women, strong measurement invariance is a precondition for being able to interpret most findings. The results of the current study demonstrate that strong gender invariance does not hold for all items of the career satisfaction scale. Implications of these findings in terms of optimal measurement of the subjective career success construct are spelled out.

Keywords: *measurement invariance; career success; career satisfaction; gender differences*

Response behavior and the career satisfaction scale: Testing for gender invariance

During recent years, career success has often been studied in relationship to a number of variables such as personality, networking and mentoring (Eby, Butts & Lockwood, 2003), socio-demographic and human capital variables (Ng, Eby, Sorensen & Feldman, 2005), motivation (Wayne, Liden, Kraimer & Graf, 1999) and gender differences (Kirchmeyer, 1998, 2002). Understanding the nature and antecedents of career success may be of interest to individuals as well as to organizations (Feldman, 1989). As contemporary careers become increasingly self-directed, individual career actors may rely on this kind of information for effectively managing their own careers. Organizations, on the other hand, might use it to rationalize their career development policies (Wayne et al., 1999). Furthermore, employees' individual success may eventually contribute to organizational success (Judge, Higgins, Thoresen & Barrick, 1999).

However, when conducting research on the topic of career success, one must first consider the issues of definition and measurement (Duxbury, Dyke & Lam, 1999; Heslin, 2005). The career success construct is commonly subdivided into two correlated yet non-interchangeable constructs: objective and subjective career success (Ng et al., 2005). Objective career success is generally measured using indicators that can be objectively evaluated by an impartial third party, such as pay, promotion and occupational status (Heslin, 2005; Judge, Cable, Boudreau & Bretz, 1995; Ng et al., 2005). Subjective career success, on the other hand, is concerned with the idiosyncratic evaluations individuals make of their own careers (Aryee, Chay & Tan, 1994; Judge et al., 1995; Melamed, 1996; Nabi, 1999).

Although, traditionally, the focus tended to be on objective career success, recent studies demonstrate an increasing emphasis on peoples' subjective evaluations of their career (Ng et al., 2005; Savickas, 1995). When asked about their subjective views on career success, career actors frequently describe managing different aspects of their career, such as maintaining a satisfactory income, finding time for their families, and pursuing new learning in a way that suggests they are thinking in terms of multiple dimensions. Furthermore, several recent publications have made apparent that career success can have a multitude of meanings, depending on the individual or subgroup constructing the meaning (Adamson, Doherty & Viney, 1998; Collin & Young, 1992). First of all, the construction of meaning (i.e. that of career and career success) is a *historically* and *culturally* specific process (Dries, Pepermans & De Kerpel, in press; Young & Collin, 2004). Furthermore, *sub-culture* and *socio-economic background* can play an important role in the way people construct the meaning of career success (Stead, 2004). Finally, and on a more individual level, developmental theories such as *socialization* and *life-cycle processes* may be relevant (Super, Savickas & Super, 1996).

However, the career satisfaction scale by Greenhaus, Parasuraman and Wormley (1990), which has been used in a manifold of studies (more than 240 Social Sciences Citation Index citations refer to research applications of the measure) and is considered "the best measure available in the literature" (Judge et al., 1995, p. 497) measures career success in a one-dimensional way (Arthur, Khapova & Wilderom, 2005). This paper will examine whether this has implications for the measurement invariance of the scale. A specific focus will be placed on gender invariance as differences between men and women in terms of their conceptualizations of what "career success" means are most frequently reported in the literature. Quite a few

studies have found that women are less likely than men to reach top levels (Cox & Nkomo, 1991; Davidson & Cooper, 1992; Schneer & Reitman, 1995; Stroh, Brett & Reilly, 1992) and earn lower incomes (Judge et al., 1995; Kirchmeyer, 2002; Schneer & Reitman, 1995; Stroh et al., 1992) but at the same time, perceive their careers to be as successful as men (Judge et al., 1995; Kirchmeyer, 1998). A meta-analytic study using 22 studies to test the impact of gender on subjective career success found that overall, although the relationship between gender and *subjective* career success was found to be non-significant, gender did predict operationalizations of *objective* career success (Ng et al., 2005).

One could wonder, then, whether men and women conceptualize the construct in the same manner (i.e. invariantly) – which is an important requirement when comparing these groups in terms of career success. As far as objective career success is concerned, this requirement of measurement invariance is obviously fulfilled. Number of promotions, or salary in dollars or euros would hold the same meaning for men as for women. With respect to subjective career success, on the other hand, the situation is somewhat more complex. Items belonging to rating scales that assess subjective career success - the career satisfaction scale by Greenhaus et al. (1990) being the most prominently used example - measure a non-observable, latent construct. Measurement invariance, for this type of scale, requires that the mathematical functions linking the items of the rating scale to the latent construct are invariant across respondent groups (Borsboom, 2006). In terms of gender, this means that men and women should interpret and use the items of the career satisfaction scale in the same way.

However, based on the literature, we might expect men and women to exhibit a different response behavior, on the different levels of invariance. First of all, since

men and women seem to define the career success construct differently (Dyke & Murphy, 2006; Parker & Chusmir, 1992; Powell & Mainiero, 1992; Russo, Kelly & Deacon, 1991; Sturges, 1999), it is well possible that they would use different frames of reference when rating themselves on the construct (i.e. *configural* non-invariance). Furthermore, it is possible that they use the scale intervals differently (i.e. *metric* invariance). Technically this means that for a given change on the latent variable, the change in the observed variables is different for both groups. In practice non-invariance regarding the scale intervals may denote two things. The first one is that the response function, mapping the score of the latent variable onto the observed variables, is gender dependent; thus that women and men use the response scale in itself differently. Second and more probable is that men and women value the respective dimensions of subjective career success differently. Since the literature indicates that gender differences in the meaning of career success exist even when occupational attainments are similar (Dyke & Murphy, 2006), metric invariance or the invariance of scale intervals can be expected. Finally, it is possible that men and women have different intercepts or different subjective “null points” when it comes to career success (i.e. *scalar* non-invariance). In this paper, we will specifically test for these assumptions. In doing so, we wish to encourage critical reflection on the way in which the career success construct is defined and measured in empirical research, a point on which we will elaborate in the Discussion

Method

Participants

Data from three large-scale survey studies were combined; each of these studies were carried out in Belgium, within the Dutch-speaking managerial

population. The inclusion criterion was being a manager, which was operationalized in each of the three studies as the combination of “having a functional title that includes the term ‘manager’” *and* “giving direction to at least one subordinate”. The total aggregated sample consisted of 596 male and 504 female managers. A description of the sample in terms of age, educational level and employment sector is given in Table 1.

INSERT TABLE 1 ABOUT HERE

As can be seen in Table 1, the distributions of both sexes across age, educational level and employment sector are very similar.

Measure

The dependent variable, subjective career success, was measured using the career satisfaction scale of Greenhaus et al. (1990). The participating managers were instructed to indicate to what extent they agreed or disagreed with each of the following statements:

1. I am satisfied with the success I have achieved in my career
2. I am satisfied with the progress I have made towards meeting my overall career goals
3. I am satisfied with the progress I have made towards meeting my goals for income
4. I am satisfied with the progress I have made towards meeting my goals for advancement
5. I am satisfied with the progress I have made towards meeting my goals for the development of new skills

Each of the preceding items is scored on a category rating scale with the following labels: “strongly disagree”, “disagree to some extent”, “uncertain”, “agree to some extent”, and “strongly agree”. All items are considered indicators of one underlying

factor, i.e. subjective career success. From now on, we will thus refer to this career success model as the one-factor model of Greenhaus et al. (1990).

Results

Multivariate normality

The current study uses the multigroup confirmatory factor analysis model since it is considered the most powerful and versatile approach for testing measurement invariance (Steenkamp & Baumgartner, 1998). As tests of equality of covariance matrices are heavily affected by deviations from multivariate normality (DeCarlo, 1997; Kline, 2005), preliminary tests were run checking for multivariate normality of the data using the SPSS macro described in DeCarlo (1997). An omnibus test for multivariate normality based on Small's statistic (DeCarlo, 1997) showed that the distributions for men ($\chi^2(10) = 245.88, p < .001$) as well as for women ($\chi^2(10) = 128.25, p < .001$) deviated significantly from the normal distribution. Therefore robust maximum likelihood was used as the estimation method for fitting the models. This estimation method analyses the data using maximum likelihood and robust standard errors. The value of χ^2 is adjusted by an amount that reflects the magnitude of observed kurtosis, which is known as the Satorra-Bentler χ^2 (Kline, 2005).

Measurement invariance

The following analyses are in accordance with the succession of tests on measurement invariance as described in the review article of Vandenberg and Lance (2000). Before the series of nested models was tested, we evaluated whether the one-factor model of Greenhaus et al. (1990) fitted the data well for each group separately (Byrne, Shavelson & Muthén, 1989). To evaluate the model fit, we took several

indicators into account, each referring to the ability of the model to reproduce the observed covariance matrix (Vandenberg & Lance, 2000). First of all, the Satorra-Bentler χ^2 is reported because of its widespread use in reports of SEM analyses. Theoretically the Satorra-Bentler χ^2 of a well-fitting model should be statistically non-significant (Kline, 2005; Vandenberg & Lance, 2000). However, it is known that the χ^2_{SB} is sensitive to minor deviations from the conceptual model and that χ^2_{SB} is highly affected by sample size (Kline, 2005; Vandenberg & Lance, 2000). Therefore it might be expected that the χ^2_{SB} is statistically significant even if the model fits well. Because of this reason we also included other fit indices such as the root mean square error of approximation (*RMSEA*), which is a badness-of-fit measure. The upper critical value for the *RMSEA* is .10, with values lower than .08 suggesting a reasonable error of approximation (Kline, 2005; Vandenberg & Lance, 2000). Finally, three goodness-of-fit indices are reported; an incremental fit index, i.e. the Comparative Fit Index (*CFI*), and two absolute fit indices, i.e. Gamma Hat (Steiger, 1989) and McDonald's (1989) Non-Centrality Index (*NCI*). For a well-fitting model these indices should exceed the critical value of .90 (Kline, 2005; Woehr, Arciniega & Lim, 2007). It is demonstrated in Table 2 that, considering the combination of all these fit indices, the one-factor model of Greenhaus et al. (1990) fitted the data well for both sexes.

INSERT TABLE 2 ABOUT HERE

In this first step it was demonstrated that the one-factor model of Greenhaus et al. (1990) is an appropriate representation of the data for both sexes; the second step proceeded with the tests on measurement invariance as described by Vandenberg and Lance (2000).

Configural invariance. As a first test, the invariance of the structure of the measurement model across groups was evaluated (Bollen, 1989; Cheung & Rensvold, 2002; Mullen, 1995; Vandenberg & Lance, 2000). If the structure of the measurement model is the same in all groups, the model is said to be *configural invariant* or *weak factorial invariant* (Horn & McArdle, 1992) and this denotes that the respondents use similar frames of reference when rating the construct (Lievens, Anseel, Harris & Eisenberg, 2007; Vandenberg & Lance, 2000). Configural invariance is tested by specifying a multigroup confirmatory factor analysis model where the one-factor model of Greenhaus et al. (1990) is fitted for both sexes. The difference with the previous analysis is that the model is no longer fitted separately for both groups but for both sexes jointly. Hence this model serves as a baseline model for the subsequent, more restrictive, measurement invariance steps. All statistical indices supported a configural invariant model. The Satorra-Bentler χ^2 was small but significant ($\chi^2_{SB}(10) = 27.53; p = .002$). The badness-of-fit measure indicated that the degree of error in the estimation of the model was rather small ($RMSEA = .057$), and finally, the three goodness-of-fit measures equaled 1 for the *CFI*, .9937 for *Gamma Hat* and .9921 for the *NCI*, which is very satisfactory.

Metric invariance. The next step proceeded with a test on *metric invariance*. Metric invariance tests whether the factor loadings are equal across groups or whether the intervals on the measurement scales are calibrated in similar ways by respondents in the different groups (Lievens et al., 2007). Therefore, with a metric invariant model, the scaling units are invariant across groups (Vandenberg & Lance, 2000). In addition to the similarity of the one-factor model of Greenhaus et al. (1990) across

gender, the constraint of equal factor loadings is added to the multigroup confirmatory factor analysis model. Since this model is nested in the previous one, it is possible to evaluate the decrease in model fit. The traditional way to test whether the model deteriorates is by performing the χ^2 -difference test or $\Delta\chi^2$ (Bollen, 1989; Cheung & Rensvold, 2002). However, since $\Delta\chi^2$ is susceptible to sample size and non-normality, Cheung and Rensvold (2002) suggested to use ΔCFI , $\Delta \textit{Gamma Hat}$ and ΔNCI as indicators of the decrease in model fit. The critical values for the difference between the fit indices of two nested models are .01 for ΔCFI , .02 for ΔNCI , and .001 for $\Delta \textit{Gamma Hat}$ (Cheung & Rensvold, 2002). The rationale for the selection of these fit indices is their independence of model complexity and sample size. Moreover, these indices are not correlated with the overall fit measures (Cheung & Rensvold, 2002). However, recently it was argued that the critical value of .001 for the $\Delta \textit{Gamma Hat}$ may be overly strict since it is sensitive to small differences in factor loadings ($<.1$), while the ΔCFI may be too insensitive (Meade, Johnson & Braddy, 2006). Therefore, these authors suggest to report the change in all fit indices; however, when the resulting findings are ambiguous, ΔNCI seems to perform optimally and should be preferred (Meade et al., 2006). When evaluating the decrease in model fit caused by equating the factor loadings, it was found that all fit indices indicated that the one-factor model of Greenhaus et al. (1990) deteriorated considerably. The Satorra-Bentler χ^2 raised from $\chi^2_{SB}(10) = 27.53$ to $\chi^2_{SB}(15) = 156.28$ while the *RMSEA* increased to .13. Moreover, the test statistics used to compare the nested models, ΔCFI , $\Delta \textit{Gamma Hat}$ and ΔNCI equalled .03, .0426 and .0543 respectively. This means that at least for one item of the career satisfaction scale, the factor loadings differed for men and women.

In a next step, the non-invariant factor loadings were identified, which was done according to the procedure outlined by Cheung and Rensvold (1999). This procedure comes down to a series of invariance tests where with each new test a different factor loading is constrained across groups. Thus in each test the factor loading of one item is constrained while all other factor loadings may vary. The factor loadings for each item in both groups, together with the decrease in *CFI*, *NCI* and Δ *Gamma Hat* when equating the factor loadings of that respective item across groups are shown in Table 3.

INSERT TABLE 3 ABOUT HERE

Table 3 shows that the largest absolute difference in factor loadings is located in item 2. The invariance of this item's factor loadings is also statistically confirmed because the three fit indices exceeded their respective critical values. Except for item 5, where all three indices were below their cut-off values, Δ *Gamma Hat* was too high for all other items too. However, we already referred to the high sensitivity of this index and because the ΔCFI and the ΔNCI did not exceed .01 and .02 respectively, we chose to retain the other four items in our analyses. Subsequently, the multigroup confirmatory factor analysis model was fitted with equality constraints on all factor loadings except for the factor loadings of item 2. The fit indices were considerably better; the Satorra-Bentler χ^2 equals 66.52 for 14 degrees of freedom while the *RMSEA* raised to .083, which still below the upper critical value. The goodness-of-fit measures used to evaluate the decrease in model fit were also satisfactory since *CFI* equals .99 ($\Delta CFI = .01$), *NCI* equaled .9764 ($\Delta NCI = .0157$), and *Gamma Hat* equaled .9813 (Δ *Gamma Hat* = .0124). As we discussed earlier, since Δ *Gamma Hat*

is very sensitive, we rely on the combination of the three measures, concluding that this model is partial metric invariance.

Scalar invariance. Now that partial metric invariance was demonstrated, the study proceeded with an invariance test on the item intercepts. A model with invariant item intercepts is said to be *scalar invariant* (Meredith, 1993; Vandenberg & Lance, 2000; Cheung & Rensvold, 2002). Scalar invariance denotes that the ratings on the observed variables are equal across groups when the score on the latent variable is zero (Cheung & Rensvold, 2002). The combination of metric and scalar invariance, called *strong factorial invariance* (Horn & McArdle, 1992; Mullen, 1995), is a requirement for the comparison of latent means since this implies that the measurements have the same intervals and null points (Cheung & Rensvold, 2002; Meredith, 1993). Hence, if strong factorial invariance is satisfied, the differences in means of the observed scale are a consequence of differences on the latent variable (Steenkamp & Baumgartner, 1998). The test on scalar invariance is performed by constraining the invariant factor loadings as well as the item intercepts to be equal across sex and by evaluating the decrease in model fit, as compared to the partial metric invariant model. All fit indices indicated that scalar invariance did not hold. The Satorra-Bentler χ^2 increased from $\chi^2_{SB}(14) = 66.52$ to $\chi^2_{SB}(19) = 379.15$ while the *RMSEA* increased to an unacceptable value of .19. Also the goodness-of-fit measures declined substantially; the *NCI* dropped to .8490 ($\Delta NCI = .1274$) while the *Gamma Hat* decreased to .8842 ($\Delta \text{Gamma Hat} = .0971$). Moreover, ΔCFI equaled .08 ($CFI = .91$). Parallel to the analyses for metric invariance, the study proceeded with tests on partial scalar invariance. Again, each item intercept was equated across gender while

all other item intercepts were allowed to vary freely. This procedure identifies the non-invariant items regarding the item intercepts.

INSERT TABLE 4 ABOUT HERE

As can be seen in Table 4, the intercepts for items 2 and 4 were not invariant across gender. For the other items, i.e. 1, 3 and 5, the model fit deteriorated when equating their intercepts across groups but the decrease in model fit was acceptable. As in the previous analyses, ΔG_{HAT} exceeded its critical value of .001 but was disregarded because of its sensitivity to very small deviations. Subsequently, the items with invariant factor loadings and invariant item intercepts were tested in one model to test for partial scalar invariance. As can be seen in Table 5, this model deteriorated too much when compared with the partial metric invariant model. Therefore, two models were fitted where intercepts of item 3 and item 5 were alternately allowed to vary free while the intercepts of the other item (5 or 3 respectively) and item 1 were constrained across gender. The fit indices in Table 5 show that neither model fitted adequately. First of all ΔNCI , the most important fit index for diagnosing measurement invariance (Meade et al., 2006), exceeded its critical value of .02. Moreover, the *RMSEA* exceeded .10, which is its upper value for a good fitting model. Ultimately, we fitted a model where only the intercept of item 1 was constrained across both sexes while the intercepts of the other items are freed. As can be seen in Table 5, this model fitted fairly well and did not deteriorate substantially from the partial metric invariant model.

INSERT TABLE 5 ABOUT HERE

In their review article, Vandenberg and Lance (2000) also mentioned a last measurement invariance test: *invariance of the error variances*. However, one can only proceed with more restrictive tests if the model fits for less restrictive test. In our case, only one item intercept is invariant across groups and therefore testing error variance invariance has little sense. Moreover, error variance invariance is a very stringent test of measurement invariance and is unnecessary for the comparison of means (Yoo, 2002). As such, this test was not performed.

Discussion

Our results show that the career satisfaction scale constructed by Greenhaus et al. (1990) does not demonstrate strong factorial invariance across gender. As described elaborately in the Results section, strong factorial invariance comprises the combination of configural, metric and scalar invariance; and its testing is a necessary prerequisite for being able to conduct meaningful substantive tests across groups (e.g. men and women). In case of the career satisfaction scale (Greenhaus et al., 1990), only one of the five items (i.e. “I am satisfied with the success I have achieved in my career”) shows strong factorial invariance. The psychological relevance of these findings, as well as the implications they hold for measurement of the subjective career success construct are discussed below.

Psychological relevance

Each of the three components of strong factorial invariance, i.e. configural, metric and scalar invariance, hold different psychological implications.

Configural invariance. In a first step of analysis, the career satisfaction scale of Greenhaus et al. (1990) was found to demonstrate configural invariance across

gender. This finding implies that the one-factor model as proposed by the authors – in which all five items measure subjective career success – holds both for men and women. In psychological terms, this implies that both sexes build on the same conceptualization of subjective career success when responding to the items of the career satisfaction scale (Lievens et al., 2007; Vandenberg & Lance, 2000).

Metric invariance. In a second step, the scale was found to possess metric gender invariance, except for the item “I am satisfied with the progress I have made towards meeting my overall career goals”. This finding indicates that the factor loadings, which represent the expected changes in the observed scores on the items for a given change on the underlying (latent) variable, are equal across gender for four of the five items. Thus, for these four invariant items the intervals used on the scale are calibrated in the same way by men and women (Lievens et al., 2007). A substantive interpretation for this finding may be that the relative importance of the invariant domains in the overall feeling of career success, as conceptualized by the different items of the career satisfaction scale, is the same for men and women. When interpreting the only non-invariant item, it seems that satisfaction with meeting the overall career goals is more important and consequently contributes more to overall career satisfaction for men than for women.

Scalar invariance. In the final step of analysis, only one item, “I am satisfied with the success I have achieved in my career”, was found to demonstrate scalar invariance. Tests on scalar invariance serve to verify whether there is invariance for each item intercept, in this case, whether men and women assign the same score to an item if their position on the underlying variable (i.e. career satisfaction) equals zero (i.e. not at all satisfied). As such, this test assesses whether there is consistency between the differences in latent means and the differences in observed means

(Steenkamp & Baumgartner, 1998). Differences in intercepts can occur because of an upward or downward measurement bias in the domain tapped by the specific item. Applied to our study the differences in intercepts mean that even when they are equally satisfied with their career, men give higher ratings on almost all items of the career satisfaction scale than women.

Summarized, when filling out the career satisfaction scale by Greenhaus et al. (1990), both sexes seem to agree on the conceptualization of career satisfaction as being a resultant of the items of the career satisfaction scale. However, the importance of these different items is not sex-invariant since satisfaction with meeting the overall career goals contributes to a stronger degree to overall career satisfaction for men than for women. Finally, men give higher ratings on almost all items of the career satisfaction scale than women, even when they feel equally satisfied with their career.

Implications for measurement of the subjective career success construct

As we have discussed in the Introduction section, gender is a specific predictor of interest in many studies on career success (Ng et al., 2005). Important implications, mostly about “glass ceilings” and discriminatory career opportunity structures within organizations on a micro-level and societies on a macro-level, are often drawn from these studies (Davidson & Cooper, 1992). However, when comparing subjective career success (i.e. career satisfaction, which is measured as a latent variable) between men and women, strong measurement invariance is a precondition for being able to interpret most findings. When comparing the two sexes by means of the career satisfaction scale of Greenhaus et al. (1990), the most frequently used measure in the literature (Judge et al., 1995), one is obviously interested in finding “real” differences in terms of their subjective career satisfaction. However, without strong measurement

invariance, observed differences on the career satisfaction scale (Greenhaus et al., 1990) may be caused by a combination of differences in the latent variable and differences in the *psychometric response behavior* exhibited by different groups of respondents (e.g. men versus women).

Our results demonstrate that strong measurement invariance does not hold for all items of the career satisfaction scale (Greenhaus et al., 1990). What are, then, the implications of this lack of strong measurement invariance? As full measurement invariance, i.e. invariance of all items of a scale, is not often encountered in practice, several researchers have suggested to relax the conditions necessary to compare differences between groups in a meaningful manner (Byrne et al., 1989; Cheung & Rensvold, 1999; Millsap & Kwok, 2004; Steenkamp & Baumgartner, 1998; Yoo, 2002). Since, in our findings, only part of the items of the career satisfaction scale (Greenhaus et al., 1990) are invariant, a relaxation could be to only consider “partial” measurement invariance. In general, four different strategies are available for dealing with partial measurement invariance (Poortinga, 1989), all of which are discussed below.

The first and most conservative strategy is to consider partial measurement invariance as a finding that renders substantive comparisons between groups impossible. The reasoning behind this strategy is that partial measurement invariance implies that the different groups rate different constructs when using the same scale (Millsap & Kwok, 2004; Poortinga, 1989). Applied to our situation, when following this strategy the conclusion is that one cannot compare subjective career success across gender when using the career satisfaction scale of Greenhaus et al. (1990). However, as such a conclusion might be overly rigid, there are three other strategies to consider.

In the second strategy, non-invariant items are omitted when computing the scale scores (Cheung & Rensvold, 1999; Millsap & Kwok, 2004; Poortinga, 1989). The problem with this strategy is that it is a-theoretical and that it implies that different items of the scale are to be used when testing different populations, thereby hindering comparisons across studies. Moreover, the usefulness of this strategy depends on the proportion of non-invariant items, which may be problematic if the original scale has few items to begin with (Cheung & Rensvold, 1999; Millsap & Kwok, 2004). For the career satisfaction scale (Greenhaus et al., 1990), this strategy is indeed problematic since only one of the initial five items demonstrates strong invariance across gender. It is, indeed, plausible that the construct validity of the scale would be affected when deleting the four other items. Given the a-theoretical nature of this strategy, and the large proportion of non-invariant items we found for the career satisfaction scale (Greenhaus et al., 1990), it appears that this strategy may also be sub-optimal.

The third alternative is somewhat different from the preceding two options since it considers the non-invariant items to be a useful source of information concerning group differences (Cheung & Rensvold, 1999; Poortinga, 1989). We will not go into detail, but the interpretation of the non-invariant factor loadings in terms of the relative importance of the different domains of career satisfaction fits in this strategy. More examples can be found in Ellis (1989), Ellis, Minsal and Becker (1989), Poortinga and Van de Vijver (1987) and Van de Vijver and Poortinga (1991).

The final strategy to deal with partial measurement invariance is to proceed with substantive comparisons even when full measurement invariance does not hold (Byrne et al., 1989; Cheung & Rensvold, 1999; Millsap & Kwok, 2004; Poortinga, 1989; Steenkamp & Baumgartner, 1998; Yoo, 2002). The problem with this strategy,

as well as with the second one, is that there is no straightforward criterion to decide how large the invariant part of the measurement model should be in order to allow meaningful comparisons (Yoo, 2002). Most researcher adhere to the opinion that only a small portion of the model may be non-invariant (Byrne et al., 1989; Cheung & Rensvold, 1998, 1999; Steenkamp & Baumgartner, 1998; Yoo, 2002). The rationale behind this argumentation is that a minority of non-invariant items will not obscure meaningful inferences from the scale (Millsap & Kwok, 2004). For our data, it is obvious that a majority of items is non-invariant, implying that the latter argumentation cannot be assumed to hold.

We must remark at this point that measurement invariance depends on the interplay of both sample and survey, and is therefore not a quality of a certain measurement instrument. As such, we cannot straightforwardly generalize our results to other studies in which gender differences regarding subjective career success are analyzed. However, we would like to conclude with some general implications for measurement of the subjective career success construct.

It is important to note that the one invariant item that was found in our results is the only general statement in the career satisfaction scale (Greenhaus et al., 1990), i.e. “I am satisfied with the success I have achieved in my career”. There are some conceivable arguments for measuring subjective career success with only one (broad) item. As was mentioned in the Introduction, the career satisfaction scale constructed by Greenhaus et al. (1990) has been criticized for its lack of broadness (Arthur et al., 2005; Heslin, 2005). However, it is precisely *because of* the generality of the statements in this scale that it can be considered as being widely applicable – a feature that would disappear when one would start composing much longer, multidimensional subjective measures of career success. Further research is necessary to evaluate how

new career measures can be developed that are both *inclusive* and *workable*. One suggestion is to work with weighted scales, in which “all possible” definitions of career success are incorporated, and respondents must indicate both the applicability of each item to their own career, as well as the their perceived relevance in light of their personal definition of career success. With such a design, both differences in career success (as measured by the weighted or unweighted applicability of the constructs to respondents’ careers) and differences in career success definitions (as measured by the evaluated relevance of each construct) could be compared between groups (e.g. men versus women). However, in order to be truly inclusive, such a subjective measure would have to contain the option of respondents adding more constructs to the scale while they are answering to it, should they feel that not all relevant constructs have been surveyed. It is clear that this would cause some methodological intricacies; perhaps using a general one-item measure of subjective career success might turn out to be the most workable solution in the end – the downside of this approach being that the dimensions used to come to one general score of subjective career success would represent a “black box” to the researchers.

In any case, it is important that careers researchers reflect thoroughly about the implications different operationalizations of career success may have on subsequent research findings since it has been repeatedly demonstrated that different measures of career success (i.e. objective versus subjective measures, one-item measures versus multiple-item measures) yield different, sometimes opposing conclusions (Arthur et al., 2005; Boudreau et al., 2001; Judge et al., 1995; Nabi, 1999, 2001).

References

- Adamson, S.J., Doherty, N. & Viney, C. (1998). The meanings of career revisited: Implications for theory and practice. *British Journal of Management*, 9, 251-259.
- Arthur, M.B., Khapova, S.N. & Wilderom, C.P.M. (2005). Career success in a boundaryless career world. *Journal of Organizational Behavior*, 26, 177-202.
- Aryee, S., Chay, Y.W. & Tan, H.H. (1994). An examination of the antecedents of subjective career success among a managerial sample in Singapore. *Human Relations*, 47, 487-510.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.
- Borsboom, D. (2006). When does measurement invariance matter? *Medical Care*, 44, 176-181.
- Byrne, B.M., Shavelson, R.J. & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin*, 105, 456-466.
- Cheung, G.W. & Rensvold, R.B. (1998). Cross-cultural comparisons using non-invariant measurement items. *Applied Behavioral Science Review*, 6, 93-110.
- Cheung, G.W. & Rensvold, R.B. (1999). Testing factorial invariance across groups: a reconceptualization and proposed new method. *Journal of Management*, 25, 1-27.
- Cheung, G.W. & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233-255.
- Collin, A. & Young, R.A. (1992). Constructing career through narrative and context: An interpretative perspective. In R.A. Young & A. Collin (Eds.). *Interpreting*

- career: Hermeneutical studies of lives in context (pp. 1-12)*. Westport, CT: Praeger.
- Cox, T.H., Nkomo, S.M. (1991). A race and gender-group analysis of the early career experience of MBAs. *Work and Occupations*, 18, 431-446.
- DeCarlo, L.T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, 2, 292-307.
- Dries, N., Pepermans, R. & De Kerpel, E. (in press). Exploring four generations' beliefs about career: Is "satisfied" the new "successful"? *Journal of Managerial Psychology*.
- Duxbury, L., Dyke, L., & Lam, N. (1999). *Career development in the federal public service: Building a world-class workforce*. Ottawa, Canada: Treasury Board of Canada Secretariat.
- Dyke, L.S. & Murphy, S.A. (2006). How we define success: A qualitative study of what matters most to women and men. *Sex Roles - A Journal of Research*, 55, 357-372.
- Eby, L.T., Butts, M. & Lockwood, A. (2003). Predictors of success in the era of the boundaryless career. *Journal of Organizational Behavior*, 24, 689-708.
- Ellis, B.B. (1989). Differential item functioning: implications for test translations. *Journal of Applied Psychology*, 74, 912-921.
- Ellis, B.B., Minsal, B. & Becker, P. (1989). Evaluation of attitude survey translations: an investigation using item response theory. *International Journal of Psychology*, 24, 665-684.
- Feldman, D.C. (1989). Careers in organizations: Recent trends and future directions. *Journal of Management*, 15, 135-156.

- Greenhaus, J.H., Parasuraman, S. & Wormley, W.M. (1990). Effects of race on organizational experiences, job performance evaluations, and career outcomes. *Academy of Management Journal*, 33, 64-86.
- Heslin, P.A. (2005). Conceptualizing and evaluating career success. *Journal of Organizational Behavior*, 26, 113-136.
- Horn, J.L. & McArdle, J.J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18, 117-144.
- Hughes, E.C. (1937). Institutional office and the person. *American Journal of Sociology*, 43, 404-413.
- Judge, T.A., Cable, D.M., Boudreau, J.W. & Bretz, R.D.Jr. (1995). An empirical investigation of the predictors of executive career success. *Personnel Psychology*, 48, 485-519.
- Judge, T.A., Higgins, C.A., Thoresen, C.J. & Barrick M.R. (1999). The Big Five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, 52, 621-651.
- Kirchmeyer, C. (1998). Determinants of managerial career success: evidence and explanations of male/female differences. *Journal of Management*, 24, 673-693.
- Kirchmeyer, C. (2002). Gender differences in managerial careers: Yesterday, today and tomorrow. *Journal of Business Ethics*, 37, 5-24.
- Kline, R.B. (2005). *Principles and practice of structural equation modeling. Second edition*. New York: The Guilford Press.
- Lievens, F., Anseel, F., Harris, M.M. & Eisenberg, J. (2007). Measurement invariance of the Pay Satisfaction Questionnaire across three countries. *Educational and Psychological Measurement*, 67, 1042-1051.

- McDonald, R.P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification*, 6, 97-103.
- Meade, A.W., Johnson, E.C. & Braddy, P.W. (2006). *The utility of alternative fit indices in tests of measurement invariance*. Paper presented at the annual Academy of Management conference, Atlanta, GA.
- Melamed, T. (1996). Career success: an assessment of a gender-specific model. *Journal of Occupational and Organizational Psychology*, 69, 217–235.
- Millsap, R.E. & Kwok, O. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, 9, 93-115.
- Mullen, M. (1995). Diagnosing measurement equivalence in cross-national research. *Journal of International Business Studies*, 3, 573-596.
- Nabi, G. (1999). An investigation into the differential profile of predictors of objective and subjective career success. *Career Development International*, 4, 212-224.
- Ng, T.W.H., Eby, L.T., Sorensen, K.L. & Feldman, D.C. (2005). Predictors of objective and subjective career success: a meta-analysis. *Personnel Psychology*, 58, 367-408.
- Parker B. & Cusmir L.H. (1990). A generational and sex-based view of managerial work values. *Psychological Reports*, 66, 947-951.
- Poortinga, Y.H. (1989). Equivalence of cross-cultural data: an overview of basic issues. *International Journal of Psychology*, 24, 737-756.
- Poortinga, Y.H. & Van de Vijver, F.J.R. (1987). Explaining cross-cultural differences: bias analysis and beyond. *Journal of Cross-Cultural Psychology*, 18, 259-282.

- Powell, G.N. & Mainiero, L.A. (1992). Cross-currents in the river of time: conceptualizing the complexities of women's careers. *Journal of Management*, 18, 215-237.
- Russo, N., Kelly, M. & Deacon, M. (1991). Gender and sex related attribution: beyond individualistic conceptions of achievement. *Sex Roles*, 25, 331-350.
- Savickas, M. (1995). Current theoretical issues in vocational psychology: convergence, divergence, and schism. In W.B. Walsh & S.H. Osipow (Eds.), *Handbook of Vocational Psychology: theory, research and practice (2nd Edition, pp. 1-34)*, Mahwah, NJ : Lawrence Erlbaum Associates.
- Stead, G.B. (2004). Culture and career psychology: A social constructionist perspective. *Journal of Vocational Behavior*, 64, 389-406.
- Steenkamp, J.E.M. & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78-90.
- Steiger, J.H. (1989). *EzPATH: Causal modeling*. Evanston, IL: SYSTAT.
- Schneer, J. A. & Reitman, F. (1995). The impact of gender as managerial careers unfold. *Journal of Vocational Behavior*, 47, 209–315.
- Stroh, L. K., Brett, J. M. & Reilly, A. H. (1992). All the right stuff: A comparison of female and male managers' career progression. *Journal of Applied Psychology*, 77, 251–260.
- Sturges, J. (1999). What it means to succeed: Personal conceptions of career success held by male and female managers at different ages. *British Journal of Management*, 10, 239-252.
- Super, D. E., Savickas, M. L. and Super, C. M. (1996). "The life-span, life-space approach to careers", in Brown, D. & Brooks, L. (Eds), *Career choice and development*, Jossey-Bass, San Francisco, pp. 121-78.

- Vandenberg, R.J. & Lance, C.E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4-70.
- Van de Vijver, F.J.R. & Poortinga, Y.H. (1991). Testing across cultures. In R.K. Hambleton & J. Zaal (Eds.), *New developments in testing: theory and applications*. (pp. 277-308). Dordrecht: Kluwer.
- Wayne, S.J., Liden, R.C., Kraimer, M.L. & Graf, I.K. (1999). The role of human capital, motivation and supervisor sponsorship in predicting career success. *Journal of Organizational Behavior*, 20, 577-595.
- Woehr, D.J., Arciniega, L.M. & Lim, D.H. (2007). Examining work ethics cross populations. A comparison of the multidimensional work ethic profile across three diverse cultures. *Educational and Psychological Measurement*, 67, 154-168.
- Young, R.A. & Collin, A. (2004). Introduction: Constructivism and social constructionism in the career field. *Journal of Vocational Behavior*, 64, 373-388.
- Yoo, B. (2002). Cross-group comparisons: a cautionary note. *Psychology & Marketing*, 19, 357-368.

Table 1: Description of the sample in terms of age, educational level and employment sector

		male	female
Age	-26	9.6%	6.9%
	26 – 35	19.6%	24.2%
	36 – 45	26.8%	36.7%
	+45	35.9%	22.6%
	missing	8.1%	9.6%
Educational level	grade school	0.2%	0.2%
	high school	7.0%	3.8%
	bachelor	28.4%	30.6%
	master	61.7%	63.9%
	missing	2.7%	1.5%
Employment sector	profit	70.6%	73.0%
	non-profit	22.1%	21.6%
	NGO	6.9%	5.3%
	missing	0.4%	0.1%

Table 2: fit indices for the one-factor model of Greenhaus et al. (1990) for both sexes

	males ($N = 569$)	females ($N = 504$)
χ^2_{SB}	$\chi^2_{SB}(5) = 18.49; p = .002$	$\chi^2_{SB}(5) = 10.67; p = .058$
<i>CFI</i>	.99	1
<i>Gamma Hat</i>	.9910	.9955
<i>NCI</i>	.9887	.9944
<i>RMSEA</i>	.067	.047
<i>90% CI RMSEA</i>	(.036 ; .010)	(.000 ; .087)

Table 3: factor loadings for men and women as well as the change in *CFI*, *NCI* and *Gamma Hat* for all items of the career satisfaction scale

item	men (<i>N</i> = 596)	women (<i>N</i> = 504)	ΔCFI	ΔNCI	$\Delta \textit{Gamma Hat}$
1	2.12	1.45	.00	.0061	.0048
2	2.62	1.00	.02	.0397	.0312
3	1.46	.99	.00	.0034	.0027
4	2.16	1.02	.00	.0137	.0108
5	.76	.54	.00	.0013	.0009

Table 4: item intercepts for men and women as well as the change in *CFI*, *NCI* and *Gamma Hat* for all items of the career satisfaction scale

item	men (<i>N</i> = 596)	women (<i>N</i> = 504)	ΔCFI	ΔNCI	$\Delta \textit{Gamma Hat}$
1	5.51	4.14	.01	.0151	.0119
2	6.04	2.81	.05	.0746	.0577
3	3.94	2.68	.01	.0174	.0137
4	5.40	2.83	.03	.0479	.0373
5	3.19	2.43	.01	.0165	.0130

Table 5: item intercepts for men and women as well as the change in *CFI*, *NCI* and *Gamma Hat* for all items of the career satisfaction scale

invariant item	χ^2_{SB}	df	$RMSEA$	ΔCFI	ΔNCI	$\Delta \textit{Gamma Hat}$
1, 3, 5	125.01	17	.11	.02	.0243	.0191
1, 3	114.68	16	.11	.01	.0203	.0159
1, 5	115.10	16	.11	.01	.0204	.0161
1	101.91	15	.10	.01	.0151	.0119