Beyond the Big Five: Personality and job performance in an e-commerce firm

Lennart Sjöberg
Stockholm School of Economics

Abstract

The purpose of this paper was to investigate of the relationship between personality and job performance with employees of a rapidly expanding Swedish e-commerce firm. Two studies were carried out with the *UPP* test of personality. In Study 1, seventy persons took the test on an Internet site. Three groups at NN AB participated: a group of top performers, selected by management, a group of employees selected at random, and 8 members of NN AB's top management. In Study 2, the validity of the *UPP* test was investigated with relation to supervisor assessments in a group of employees in customer service at NN AB. In Study 1, large and significant differences were found between the group of employees selected at random, and the other two groups. In terms of correlations, validity was = 0.44. In Study 2, the tested subjects were assessed by supervisors using a comprehensive assessment form (40 variables), which could reduced to three criterion dimensions. The validities for the three criterion dimensions of value for the company (halo), efficiency and social functioning were 0.66, 0.52 and 0.40. Analysis of proxy criteria (work motivation and similar attitude scales) gave similar results

Key words: Job achievement, personality, test validity, halo

A large number of studies have been carried out in order to investigate personality and intellectual ability in relation to job performance. The purpose of such studies has often been to identify efficient predictors of job performance, to be used for selection among job applicants. Some users of personality tests have been quite optimistic about developments of such tests since about 1990. Earlier work showed only weak relationships between personality and various criteria of job performance (Ghiselli, 1966).

However, several meta analyses have been carried out and have shown that job performance is weakly related to personality, in particular as measured with the Five-Factor Model (FFM), the "Big Five" personality factors (Barrick, Mount, & Judge, 2001; Goodstein & Lanyon, 1999; Salgado, 2003; Schmidt, Shaffer, & Oh, 2008). In terms of correlations, the results obtained have sometimes been around 0.3 for the correlation between a personality variables (agreeableness and conscientiousnessssss) and job performance, more often <0.25[1].

Most of these studies were performed before the breakthrough of the Internet. The current situation changes rapidly in the sense that e-commerce takes over more and more of commercial activity. It is important to investigate if previous findings on personality and job performance hold true even in the context of new IT technology. These are typically young companies, some of which have expanded vary rapidly; and this is certainly the case with the company investigated here. It started less than 10 years ago and has now over 700 employees. How well do the "Big Five" factors work in such a context, and are new concepts called for?

An important reason for the relatively weak relationships found in previous work is probably that the FFM factors are too general and too little focused on job performance. It is a general principle of behavioural data, often supported empirically, that strong relationships are found between specific to specific variables or general to general variables, and not between specific and general variables (Sjöberg, 1980, 1982).

General factors, often used in personality psychology, have not been very useful to understand behaviour. Narrow personality dimensions have been found to be more strongly correlated with job performance than broad dimensions in a number of studies (Christiansen & Robie, 2011; Conte & Gintoft, 2005; Griffith & Jenkins, 2004; Tett, Steele, & Beauregard, 2003). Some recent work going beyond the FFM has tended to be more promising, showing relatively strong correlations between personality and job performance, around 0.5 - 0.6, almost at the same level as tests of intellectual ability (Bergman, Lornudd, Sjöberg, & von Thile Schwartz, 2014).

The present paper describes two further studies of job performance and personality, based on the notion that narrow-band personality dimensions can be expected to be more strongly related to job performance in a general sense, than the FFM factors. The criterion used is that of judgments by supervisors and managers, and can therefore be expected to reflect a combination of job performance, (the reverse of) counterproductive work behaviors (CWB) and organizational citizenship behaviors (OCB). In addition, the work reported here uses a personality test - the *UPP* test - which measures several focused dimensions and which has been found, in earlier work, to correlate more strongly with job performance than tests of the

The *UPP* test

The *UPP* test was developed over a period of 8 years, and validated both with regard to similar constructs in the literature (construct validation) and external measures of job performance. It is not based on a single theory of personality, for two important reasons. First, there are many personality theories, none of them convincingly validated in full, although most of them may have some validity. Second, and more important, the *UPP* test measures dimensions close to the needs of practitioners, such as control orientation and emotional intelligence, which are not often found in all or most other personality tests. In this way, the chance is maximized to obtain relevant and valid information about the tested persons. The concepts used in the test are explained in Tables A1-A5 in Bergman et al. (Bergman et al., 2014). Sample items from the test are given in Table 1 of the Bergman et al. paper (Bergman et al., 2014). The test is oriented towards personality traits related to competencies in working life.

The *UPP* test is a self-report test. It is natural to assume that many people will respond tactically in a situation where much is at stake for them, and research has shown that they in fact do (Peterson, Griffith, & Converse, 2009; Robie, Brown, & Beaty, 2007; Sjöberg, 2012), but to different extents (McFarland & Ryan, 2000). The *UPP* test suggests a solution for this problem which is integrated in scoring the test data. Social desirability is measured in the test, both overtly and covertly. The overt measure is of the Crowne-Marlowe type (Crowne & Marlowe, 1960) while the covert measure is based on items which are not as clearly social desirability indicators, but were selected because they correlated strongly with the overt scale. Though the use of SD scales is a well-known approach, the emphasis on using models to estimate test scores with effects of faking removed seems to be less widely used. It is particularly important since different dimensions in a personality test can be expected to be differently related to measures of faking.

Reliability of the *UPP* test is about 0.7 (both internal consistency and stability over 4 weeks), construct validity is at the same level. Validity in relation to to external criteria is around 0.45 (Sjöberg, 2013).

The *UPP* test has a number of properties, among others:

- With the help of focused scales beyond the Big Five, such as emotional intelligence and positive attitude, *UPP* achieves strong correlations with relevant work attitudes, motivation and performance. Validity with regard to work performance has been found to be around 0.5. Construct validity and reliability both have a mean of 0.75.
- The test also measures the perception of the work situation, job interest and satisfaction, as well as willingness to work. These dimensions, important in themselves, correlate strongly with externally assessed job performance and can therefore be used as "proxy criteria", a fast and convenient way to get a preliminary validation of *UPP* and other tests. The procedure works only with incumbents since

data on job interest etc. are not available before a person has been hired and had some experience with the job.

- *UPP* measures mood in testing and attitude to the test in the final stage of taking it ("face validity"). Low mood may lead to misleading results and can call for re-testing. A positive attitude to the test is very important for credibility of the recruitment process of the company using the test.
- Effects of faking are eliminated with a specific, validated correction. The correction is fitted to each test dimension separately and is not based on a comprehensive assessment of a person's "credibility". See Sjöberg (Sjöberg, 2015).

Study 1

Purpose and design

This study was carried out with employees of NN AB, a rapidly expanding Swedish e-commerce company. The purpose of Stucy 1 is to give an account of a study of the personality differences among three groups of employees at NN AB:

A. A group selected by management for their superior job performance
B. A group selected from the employees at random
C. Top management

The number invited to participate, and the response rate after two reminders, are provided in Table 1.

Table 1 about here

The participants took the test on the Internet, in an unproctored situation. In addition, the participants were instructed that the results would not have any personal significance for them.

Results

Strategy of data analysis

In this paper a mostly descriptive orientation to data analysis is taken, rather than hypothesis testing. The results in all test scales were therefore first standardized to mean = 0 and standard deviation (SD) = 1, z-scores. Mean and SD were computed for the entire group. The purpose was to get a metric which has a straightforward statistical meaning, something which is not achieved with the raw scores. The means of the z-scores were computed in all three groups. Differences between groups were measured as differences between mean z-scores.

What are to be regarded as "small" and "large" differences is of course to some extent arbitrary. An often used convention is the following (Cohen, 1988):

0.2 - small effect
0.3- 0.7 - moderate effect
0.8 and larger - large effect

In terms of correlations the corresponding terminology is:

r = 0.10 to 0.29 is a weak correlation, *r* = 0.30 to 0.49 moderate, and *r* ≥ 0.50 strong. Since many tests of significance were carried out it must be expected that some of the weaker correlations or differences were due to chance.

Group differences

The analyses were performed in the way described above, and graphs were created to give a comprehensive overview of the results. Fig. 1 gives the mean differences between groups A and B, Fig. 2 between C and B and Fig. 3 between all NN AB data combined and the norms for the English version of the test[2] Note that A contains the top performers, B the randomly selected employees and C members of the top management.

Figs 1 and 2 about here

Moderate positive differences between A and B are found in

- Openness
- Job interest
- Social skills

Moderate negative differences are found in

- Agreeableness
- Perfectionism
- Work/life balance

These six scales could be combined to form a predictive index. However, this procedure risks "capitalizing on chance", and the size of the samples precludes the use of more advanced and efficient statistical methods, such as multiple regression. It was therefore preferred to create an index based on all of the test variables (transformed to a common metric of z-scores), after reversing a few scores of variables where group B scored better than group A. All variables were entered with equal weights[3]. A one-way analysis of variance, using groups (A, B and C) as independent variable gave a highly significant *F* value of 8.158, with degrees of freedom = 2,69, $p < 0.0005$. This corresponds to a correlation of 0.443 between personality as measured by the test and job performance. The means of the over-all index were:

A: 0.39
B: -0.43
C: 0.41

The differences between B and the other two groups were highly significant, while the difference A-C was not significant. Note that the differences between A and C on the one hand, and B on the other are dramatic in size (>0.8). In correlation terms, the validity of the test was quite satisfactory (>0.4) and of unusual size. Fig. 3 describes the differences between all NN AB data and the norm group.

The figure shows that NN AB data gave a more positive picture of the respondents than the English norm. Especially work motivation and work related attitudes came out as be more positive. There were some differences in impression management scales, albeit small, around 0.2.

Social desirability responses: further analysis

The test situation was not one where the test takers could expect important consequences of their test results; yet some of them may have suspected such consequences. Alternatively, the context of being an employee and requested by management to take a test may have created a set where it was seen as important to give "good" answers to the test items. To check on this possibility, means and standard deviations of the overt and covert SD scales were computed in all three groups, as well as corresponding data from two large norm data sets: applicants for management jobs in several other firms (group D, N = 371) and people who took the test just in order to be part of a norm group (group E, N = 904). The results for the five groups are given in Table 2.

The table shows a very large span between job applicants and NN AB groups, and no clear and consistent differences among the three NN AB groups. As noted above, groups A and C from NN AB differed strongly from group B. Controlling for social desirability responses, using analysis of covariance, the groups still differed significantly, $F(4,69) = 5.499$, $p<0.0005$. Hence, the group differences could not be accounted for by differences in social desirability responses.

It should be noted that job applicants gave uniquely large SD responses, which was expected. This sort of validation of the SD scales has been frequently noted, e.g. in a study of applicants to the Swedish Army officers' training (Sjöberg & Wolgers, 2012).

Big Five vs specific personality dimensions

Two indices were formed, one based on the Big Five scales, and one on the other scales. These indices were standardized to mean = 0 and standard deviations = 1 across the entire sample. Means were computed for each group, see Fig. 4.

The figure shows only small differences (non-significant) for the Big Five index [$F(2,69) = 0.289$, ns] but large and significant differences for the index based on non-Big Five scales [$F(2,69) = 4.030$, $p = 0.022$]. It is particularly interesting that top performers differed vey clearly from employees selected at random.

Discussion

A clear conclusion could be drawn: top performers and top management differed strongly from randomly selected employees in terms of personality traits. Based on our previous work with the *UPP* test, these results were expected.

The validity of the *UPP* test can be estimated to be 0.44, on the basis of the present data. It should be noted that this value if is considerably higher than meta-analytic estimates of the validity of the FFM (Schmidt et al., 2008). As noted in the introduction, more focused and narrow personality variables are more likely to be valid predictors of job performance than the general "Big Five" dimensions.

It is likely that the top performers were especially good at customer relations, suggesting that their personality profiles were indicative of good performance in that respect. Little research has been reported on this topic in e-commerce firms, but a paper by Bologna et al. (Bologna et al., 2005) suggested that Holland's system of vocational interests (Holland, 1996, 1997) applied to employees could be used for personality measurement with the aim of optimizing customer experience.

When comparing NN AB data to the English norm we found the former to give a more positive picture, especially in scales related to work motivation and attitudes. This finding could be interpreted as a sign of NN AB's efficient recruitment policies but there should be some caution here since the norm group came from a different culture and little is known about cultural differences in the test scales.

It is noteworthy and expected that excessive concern with detail (perfectionism) was a negative predictor. Perfectionists rarely finish their tasks in time and tend to be little creative. It was perhaps more unexpected that agreeableness also was a negative predictor. However, it is possible that employees who are very eager to do what they are told are perceived as less useful to the company because of a lack of own initiative and courage to propose their own solutions and viewpoints.

The relatively strong effect of financial motivation suggests that economic incentives are a powerful factor for employees in the present context, but not for top management. Work/life balance was particularly low for top performers and for top management. It is possible that they gave priority to work rather than to family life, a common finding among managers. Top performers and top management were fairly close in results on the *UPP* test, suggesting that employees who performed well were similar to management whether they hade been promoted to top management or not.

Study 2

Purpose and design

Validity of the *UPP* test has been documented previously, using proxy criteria (Sjöberg, 2008), external criteria (manager career, leadership assessed by coworkers) (Bergman et al.,

2014; Sjöberg, 2008)}, as well as in Study 1 of the present paper. However, Study 1 used a global assessment as criterion and it was considered to be important to investigate different aspects of job performance as criteria. The purpose of Study 2 was therefore to investigate validity of the test in customer service with the help of supervisors' assessments of a large range of criteria and through further study of proxy criteria.

Service quality is of great importance for this type of business (Ding, Hu, Verma, & Wardell, 2010; Groth & Goodwin, 2011), as well as in many other contexts. Personality and attitudes have previously been found to be important in customer service work (Bettencourt, Gwinner, & Meuter, 2001; Brown, Mowen, Donavan, & Licata, 2002).

Criterion validity has many aspects (Kline & Sulsky, 2009). An important question is which criterion dimensions the test should be related to. Motowidlo, Hogan et al. distinguish between core duties of the job and social functioning at work (Borman & Motowidlo, 1997); (Motowidlo, Borman, & Schmit, 1997) .

The *UPP* test measures 13 personality variables and six attitudes to work and work motivation (Sjöberg, 2010/2012) . All 19 dimensions can not be assumed to correlate with all criteria ; it would in fact be surprising if this were the case. What connections can be expected also depends on what is required in the job (J. Hogan & Holland, 2003). For example, if there are no requirements on creativity, it is not reasonable to expect that creativity should have high validity in the context.

Method

Procedure and participants

Employees within the Department of Customer Service of NN AB were invited to participate in the study, 65 in total. They were offered a fee of SEK 250, and if they so desired a test report, certificate and /or diploma. Moreover, they were promised an additional 250 SEK if 90% of those invited attended. That did not happen, 53 of 65 participated, i.e. 82%. Of these, 14 were men and 39 women, age 17-29 years, mean 21.9 years. Education: high school 33, had started college 11, graduates from college 9. They worked as customer service representatives, fraud investigators and debt collection agents.

Assessment

Managers and supervisors assessed the participants' work and social functioning on the job on a form, designed for this study which can be found on the web [4]. The form comprised 40 appraisal variables partly related to overall qualitative and quantitative aspects of work performance, as well as more specific aspects. The form was inspired by Hogan's concepts "getting along and getting ahead" (J. Hogan & Holland, 2003; R. Hogan, Roberts, Walsh, Craik, & Price, 2000).

Results

The first section describes the test characteristics of this group and the group averages, followed by a section on the structure of the criterions variables. Results of the validation analyses are then reported.

Test characteristics and group averages

Reliabilities of the test variables are summarized in a table available on the web[5]. They were consistently high and at the same level as in other studies. The values of data quality indexes are shown in Table 3 where they are compared with the norm group.

Table 3 about here

There was a tendency to fake the answers, and to affirm positive statements and reject negative statements. High values of differentiation and structural similarities suggest at the same time that the tested persons took great care to answer the test. These results suggest that the tested persons perceived the situation as if they were tested in a high-stakes situation. Since the employer would not be informed about the performance of individuals tested this may have been an unwarranted interpretation. On the other hand, they themselves shared in the results and they knew that the company would get some information about of the averages for the group.

Table 4 gives the differences in mean values of data quality between the present group and the norm.

Table 4 about here

The results show:

• Differences between the group of tested employees and normative data in personality were significant in only three of the scales: emotional stability (present group lower), openness and perfectionism (present group higher)

• Scale values in results orientation and willingness to work with changes were higher for the group of employees which also had a slightly lower work interest.

Structure of the criterion variables

A component analysis was conducted of the 40 criterion variables to provide an overview of data and simultaneously capture the dominant and important aspects. Fig. 3 shows the results of a scree test (Cattell & Vogelmann, 1977) that is used to determine how many components (factors) that provides a good, reasonable approximation to the data. The figure shows that two components give a very good approximation, explaining 53.2% of the total variance in criterion assessments.

The loadings in the two components after rotation (oblique) are shown in Table 5. Only values > 0.4 or <-0.4 have been included in the table. The components had a weak positive correlation. Two criteria were excluded because they did not have a loading in any of the components. (Does not have time to finish jobs, the number of completed tasks).

Table 5 about here

The result is interesting in several ways:

A very clear structure emerged in the data: one component measures effectiveness on core job tasks and one measures the social adaptation to the job. This agrees well with previous research on the structure of job performance criteria. Three criterion variables are loaded in both components. These variables measure the assessor's views on whether the employee is of such value to the company that they strongly want to keep him or her.

Three indices measuring efficiency, social function and "hire and retain", called value to the company (the three complex criteria) were created for further analysis. Reliabilities (Cronbach's alpha) of these criterion indices were:

Efficiency: 0.92, 10 scales
Social functioning: 0.93, 15 scales
Value to the company: 0.93, 3 scales

Efficiency and social functioning were related to the assessment of value to the company. The adjusted multiple correlation (R^2_{adj}) was = 0.78. Beta weights (standardized regression weights) were 0.49 and 0.56 for efficiency and social functioning, respectively. They were, in other words, of roughly equal importance.

Criteria in relation to test results

Table 6 gives the correlations between the test variables and the three composite criterion variables.

Table 6 about here

It is striking that the correlations are highly variable. The low values, such as for creativity, are natural considering the job content. A multiple correlation should not be calculated with N = 53 and the number of independent variables = 19, but the correlations with indices where all test variables are included with equal weights are informative: 0.33, 0.28 and 0.20.

The estimated value to the company seems to be associated primarily with the following test variables:
• Positive attitude
• Perseverance
• Emotional stability
• Extraversion
• Willingness to cooperate
• Life/work balance

These test variables were of general value also for the other criteria. A combined index with these six test variables with equal weights gave, as shown in Table 7, correlations of 0.50, 0.39 and 0.29 of the criteria value for the company, efficiency and social function. It is important to note that 3 of the 5 crucial personality factors included in the *UPP* test do not belong to he overarching dimensions of the five factor model. Performance at the job could be reasonably well predicted largely by other scales of the test.

The procedure for selecting variables may have capitalized on chance. To check this possibility, the sample was divided randomly into two subgroups and the selection of variables was done on the data from one of the groups, correlations computed in the second. These correlations are given in Table 4. They were only slightly lower than those computed on the entire sample.

The correlations between personality scales and the 6 proxy criteria an be found on the web[6]. The table shows high correlations between personality variables and proxy criteria. Positive attitude stands out as the most important dimension, followed closely by self-confidence. In only a few cases Big Five scales had better validity than the *UPP* test scales. The average correlation between personality scales and proxy criteria was computed. This average correlation was related to personality scales validities with regard to supervisors assessments. It then turned out that the average proxy validity correlated very highly with the correlations obtained with the value to the enterprise (0.64), lower but still fairly highly with efficiency (0.39) and social functioning (0.43). This result is important because it suggests that proxy validation can be expected to produce results similar to those obtained through validation against supervisor assessments, particularly for an overarching criterion such as value to the company. Fig. 5 provides a plot of the two series of correlations to each other.

Faking

The *UPP* test has an unusual advantage in the method of correction for faking used. The method has been validated (Sjöberg, 2015). It is also evident in practical applications that the effect of correction can be very important in particular cases. Table 4 shows, however, some positive correlations between measures of faking and criteria. Table 5 shows similar results for the proxy criteria. It is therefore important to examine the validity also for the corrected variables. This has been done; correlation between the two series of validities is very high, see Fig. 6.

Fig. 6 about here

The figure shows that the structure of the corrected and uncorrected variables in relation to criteria was very similar, but the corrected variables had somewhat lower validity. After squaring validities, averaging, and finally return to the same scale as the original correlations result was 0.24 for the uncorrected variables, 0.18 for the corrected ones.

Conclusions

The overall judgment of value to the company could be a halo effect driven by affective reactions (Lefkowitz, 2000). Maybe some people have a personality which makes them especially likable. This hypothesis got some support from the fact that more specific criteria could not be so well predicted by personality scales. On the other hand, it seems reasonable to hire people who will be well liked and perceived to be a great value to the firm.

The results of Study 2 also showed that

• The tested group was characterized by only a few personality differences from the norm group. The lower emotional stability may be due to demographic factors or mood
• The tested group showed higher values in results orientation and willingness to work with changes, but was slightly lower than the norm group in work interest
• Validities with regard to the criteria of value for the company, efficiency and social function gave high to medium values
• The correlations were stronger with the proxy criteria, and these correlations were very similar those obtained during validation against those of independent supervisor assessments
• Test scales after correction for faking had a slightly lower validity

Author's note

Henrik Nilheim did the IT work for both studies.

Table 1. Number of participants and response rate		
Group	Number invited	Number of respondents
A (Especially good performers)	35	31
B (Employees selected at random)	36	31
C (Top management)	8	8
Norm		127

Table 2. Means and standard deviations of Overt and Covert social desirability responses				
	Overt scale		Covert scale	
Group	Mean	Standard deviation	Mean	Standard deviation
A (NN AB, top performers)	3.27	0.51	3.85	0.40
B (NN AB, randomly selected employees)	3.80	0.52	3.78	0.40
C (NN AB, top management)	3.64	0.57	3.80	0.52
D (job applicants)	3.93	0.52	4.39	0.35
E (norm group)	3.48	0.64	3.91	0.45
Span in pooled SD units	1.20	0.55 (pooled)	1.43	0.42 (pooled)

Table 3. Differences in data quality, Study 2 (standardized values, M=0, SD= 1).	
Keystone	Difference between group of employees and norm data
Overt faking	0.18
Covert faking	0.49 **
Differentiation of responses	0.46 **
Acquiescence	0.48 **
Structural similarity	0.27

** p <0.01

Table 4. Differences between data from customer service employees and norm data, standardized scales	
Scale	Mean differences between employees' and norm data
Extraversion	0.16
Agreeableness	0.03
Emotional stability	-0.45'***
Openness	0.40***
Conscientiousness	0.12
Endurance	0.05
Willingness to cooperate	0.14
Positive attitude	-0.05
Self conficence	0.15
Social ability	0.18
Emotional intelligence (self report)	0.00
Creativity	0.09
Perfectionism	0.31**
Work satisfaction	0.09
Willingness to work	0.05
Results orientation	0.50***
Willingness to work with changes	0.41**
Work interest	-0.23
Work/life balance	0.04

* $p < 0.01$
** $p < 0.05$

Table 5. Component loadings	Component 1	Component 2
Has a good understanding of the tasks	0.94	
Overall quality of his/her work	0.86	
Is quality conscious	0.83	
Shows interest in the job	0.80	
Overall performance	0.77	
Solves problems that may arise on the job	0.76	
Is careful with important details	0.75	
Makes every effort to make a good job	0.74	
Exerts oneself	0.72	
Deserves to get a bonus	0.69	
Puts a value on achieving results on the job	0.68	
Is involved in the company's results	0.67	
Works energetically	0.67	
Plans work on a good way	0.67	
Works independently	0.66	
Shows a positive development of attitude, motivation and skills	0.60	0.40
Is the kind of person that we would be happy to employ and keep in the company	0.55	0.51
Is creative, hs good ideas	0.54	
Has the will to stay on this job	0.53	
We definitely want to keep this person in the company	0.52	0.52
Is an effective employee in our company	0.52	
Works overtime willingly	0.47	
Inefficient time use	-0..45	
Would be pleased to work with others		0.82
Extent of absence from the job on the basis of disease		-0.80
Is helpful in relation to work mates		0.74
Is a pleasant and winning person		0.72
He/she is positive and popular		0.69
Extent of absence for unknown reasons		-0.68
Is reliable		0.68
It is easy for managers to have to do with him/her		0.68
Solves social problems in the workplace, e.g. conflicts		0.68
Extent of absence for other legitimate reasons		-0.65
Has a positive attitude to the company		0.59
Receives criticism in a good way		0.59
Allows others get praise or rewards for good work		0.58
Complains about working conditions and/or salary		-0.56
Has a positive attitude to customers		0.44

Table 6. Correlations between test scales and three criterion indices.			
	Value to the company	Efficiency in core tasks	Social function
Extraversion	0.39 **	0.30 *	0.38 **
Agreeableness	0.24	0.06	0.27 *
Emotional Stability	0.40 **	0.17	0.32 *
Openness	0.17	0.08	0.08
Conscientiousness	0.04	0.05	0.12
Stamina	0.45 **	0.19	0.29 *
Willingness to cooperate	0.37 **	0.11	0.26
Positive Attitude	0.53 **	0.20	0.35 *
Self-Confidence	0.31 *	0.20	0.24
Social ability	0.14	-0.07	0.26
Emotional intelligence	0.03	-0.23	0.11
Creativity	0.17	0.09	0.09
Perfectionism	-0.17	-0.10	-0.13
Job satisfaction	0.29 *	0.12	0.15
Willingness to work	0.29 *	0.07	0.18
Results orientation	0.32 *	0.19	0.22
Willingness to work with changes	0.23	0.02	0.20
Work interest	0.33 *	0.33 *	0.11
Work/life balance	0.39 **	0.20	0.31 *
All test scales with equal weights, perfectionism reversed	0.34 **	0.28 *	0.20
Six best test scales, equal weights	0.50 **	0.39 **	0.29 *
Cross validated correlation on random half of the sample	0.49	0.28	0.32
Validity corrected for measurement errors and reduced variability	0.66	0.52	0.40
Covert faking	0.27 *	0.00	0.28 *
Overt faking	0.41 **	0.17	0.31 *
Range of correlations	-0.17 - +0.53	-0-10 - +0.33	-0.13 - +0.38
Median	0.29	0.11	0.24

$* \; p < 0.05$
$** \; p < 0.01$

Figure Captions

Figure 1. Mean differences between groups A and B, rank ordered

Figure 2. Mean differences between groups C and B, rank ordered

Figure 3. Mean differences between NN AB and Norm Data

Figure 4. Mean values of the Big Five dimensions, and for all other scales combined, for three groups.

Figure 5. Plot of validity according to the proxy criteria against validity in relation to supervisor assessments.

Figure 6. Plot of validity of the corrected test scales against the validity of the uncorrected scales.
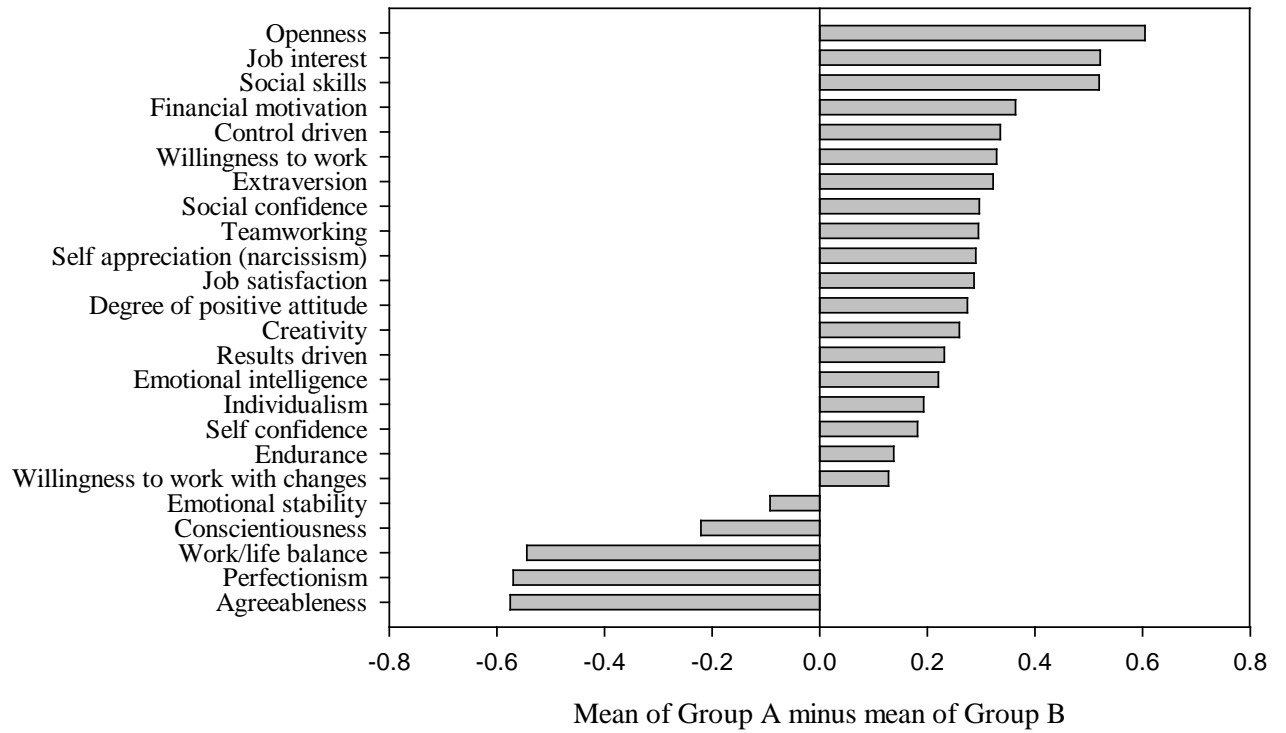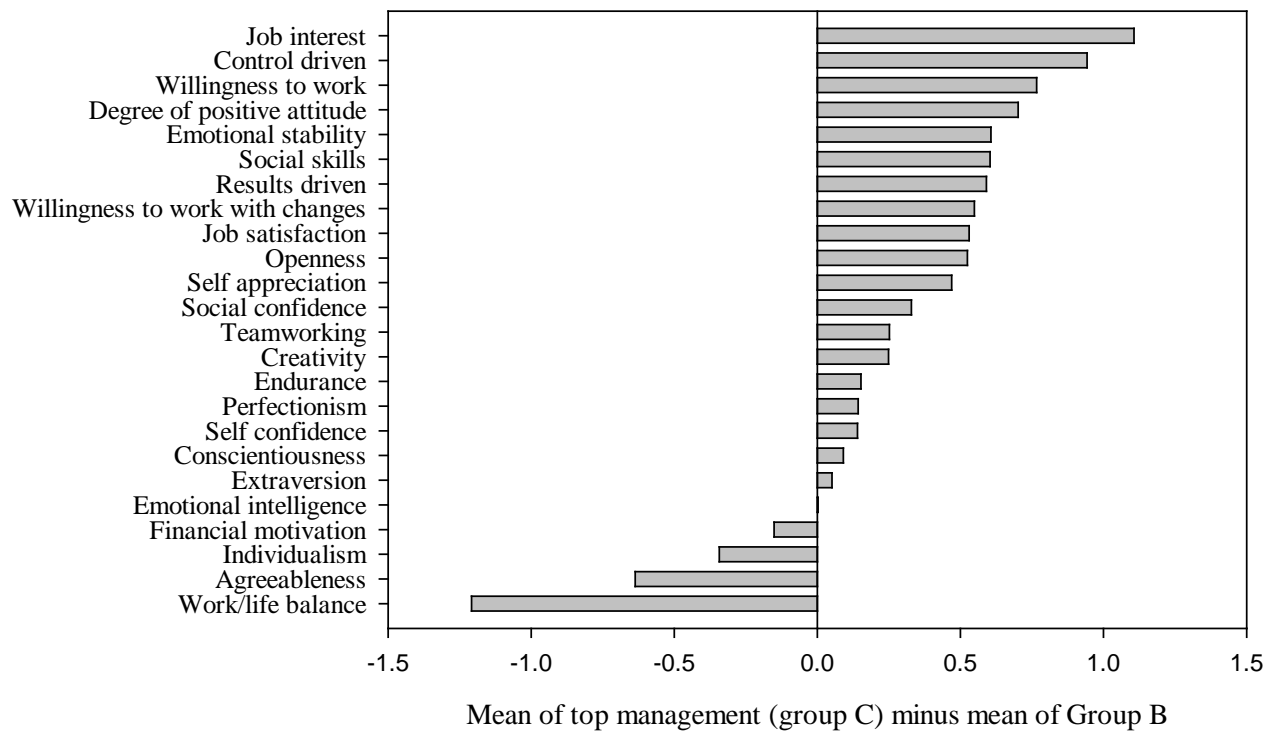
Fig. 1

Mean of Group A minus mean of Group B

Fig. 2

Mean of top management (group C) minus mean of Group B
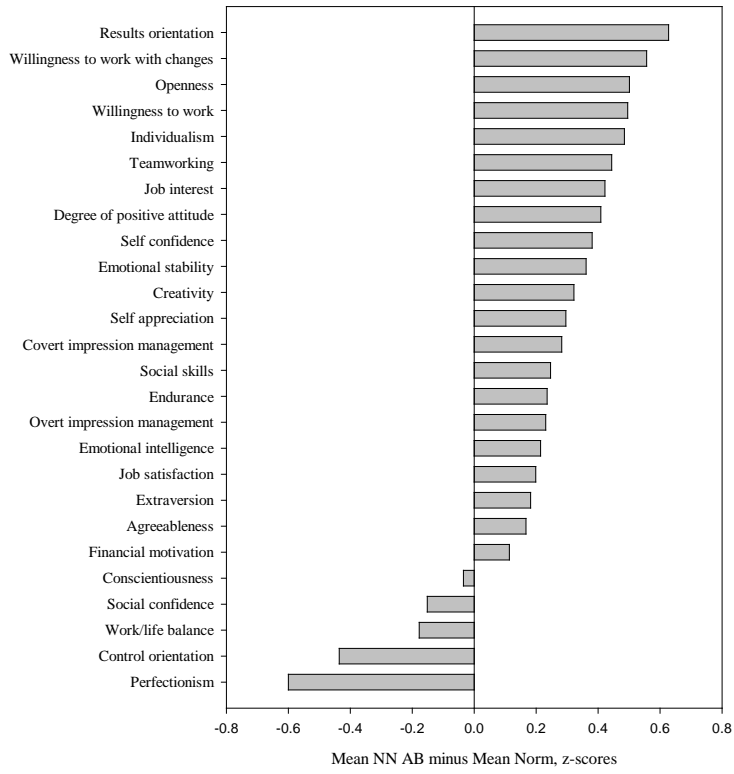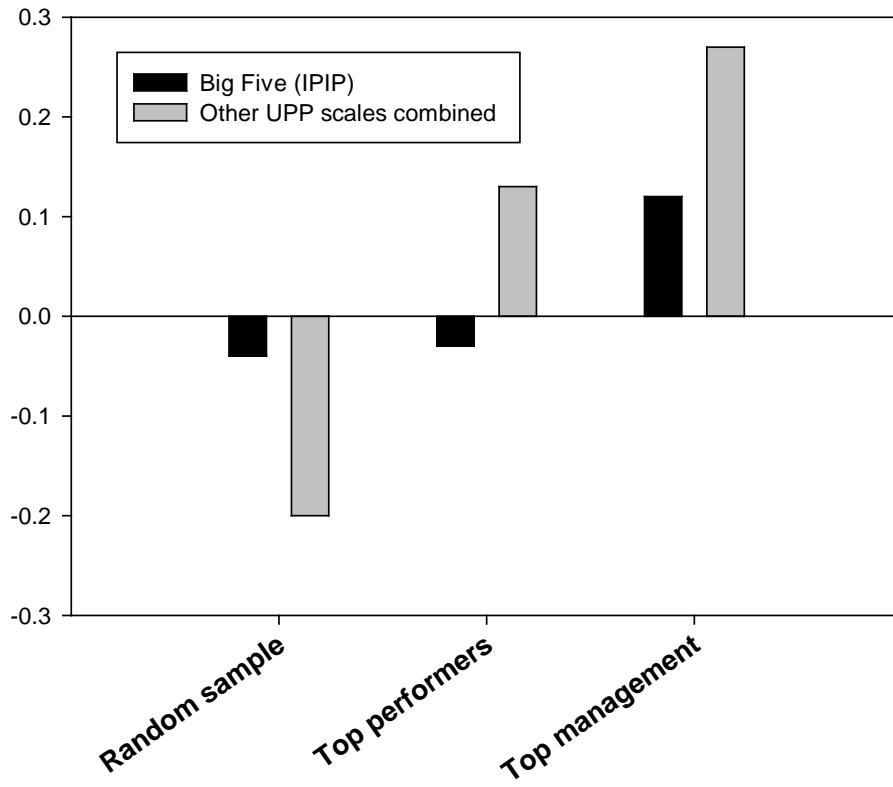
Fig. 3

Mean NN AB minus Mean Norm, z-scores

Fig. 4

Mean, standardized scores

References

Allport, G. W. (1921). Personality and character. Psychological Bulletin, 18, 441-455.

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, 9-30.

Bergman, D., Lornudd, C., Sjöberg, L., & von Thile Schwartz, U. (2014). Leader personality and 360-degree assessments of leader behaviour. Scandinavian Journal of Psychology, 55, 389-397.

Bettencourt, L. A., Gwinner, K. P., & Meuter, M. L. (2001). A comparison of attitude, personality, and knowledge predictors of service-oriented organizational citizenship behaviors. Journal of Applied Psychology, 86, 29-41.

Bobko, P., Roth, P. L., & Buster, M. A. (2007). The usefulness of unit weights in creating composite scores: A literature review, application to content validity, and meta-analysis. Organizational Research Methods, 10, 689-709.

Bologna, C., Rosa, A. C. D., Vivo, A. D., Gaeta, M., Sansonetti, G., & Viserta, V. (2005). Personality-based recommendation in e-commerce. http://ceur-ws.org/Vol-997/empire2013_paper_1.pdf.

Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. Human Performance, 10, 99-109.

Brown, T. J., Mowen, J. C., Donavan, D. T., & Licata, J. W. (2002). The customer orientation of service workers: Personality trait effects on self- and supervisor performance ratings. Journal of Marketing Research, 39, 110-119.

Cattell, R. B., & Vogelmann, S. (1977). A comprehensive trial of the scree and kg criteria for determining the number of factors. Multivariate Behavioral Research, 12, 289-325.

Christiansen, N. D., & Robie, C. (2011). Further consideration of the use of narrow trait scales. Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 43, 183-194.

Clarke, S., & Robertson, I. T. (2005). A meta-analytic review of the big five personality factors and accident involvement in occupational and non-occupational settings. Journal of Occupational and Organizational Psychology, 78, 355-376.

Cohen, J. (1988). Statistical power analysis for behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Conte, J. M., & Gintoft, J. N. (2005). Polychronicity, big five personality dimensions, and sales performance. Human Performance, 18, 427-444.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. Journal of Consulting and Clinical Psychology, 24, 349-354.

Ding, D. X., Hu, P. J.-H., Verma, R., & Wardell, D. G. (2010). The impact of service system design and flow experience on customer satisfaction in online financial services. Journal of Service Research, 13, 96-110.

Ghiselli, E. E. (1966). The validity of occupational aptitude tests. New York: Wiley.

Goodstein, L. D., & Lanyon, R. I. (1999). Applications of personality assessment to the workplace: A review. Journal of Business and Psychology, 13, 291-322.

Griffith, R., & Jenkins, M. (2004). Using personality constructs to predict performance: Narrow or broad bandwidth. Journal of Business and Psychology, 19, 255-269.

Groth, M., & Goodwin, R. E. (2011). Customer service. In Apa handbook of industrial and organizational psychology, vol 3: Maintaining, expanding, and contracting the organization. (pp. 329-357): Washington, DC, US: American Psychological Association.

Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. Journal of Applied Psychology, 88, 100-112.

Hogan, R., Roberts, B. W., Walsh, W. B., Craik, K. H., & Price, R. H. (2000). A socioanalytic perspective on person-environment interaction. In K. H. Craik & R. H. Price (Eds.), Person-environment psychology: New directions and perspectives (2nd ed.) (pp. 1-23). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Holland, J. L. (1996). Exploring careers with a typology: What we have learned and some new directions. American Psychologist, 51, 397-406.

Holland, J. L. (1997). Making vocational choices: A theory of vocational personalities and work environments (3rd ed.). Odessa, FL, US: Psychological Assessment Resources.

Kline, T. J. B., & Sulsky, L. M. (2009). Measurement and assessment issues in performance appraisal. Canadian Psychology/Psychologie canadienne, 50, 161-171.

Lefkowitz, J. (2000). The role of interpersonal affective regard in supervisory performance ratings: A literature review and proposed causal model. Journal of Occupational and Organizational Psychology, 73, 67-85.

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. Journal of Applied Psychology, 85, 812-821.

Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. Human Performance, 10, 71-83.

Peterson, M. H., Griffith, R. L., & Converse, P. D. (2009). Examining the role of applicant faking in hiring decisions: Percentage of fakers hired and hiring discrepancies in single- and multiple-predictor selection. Journa of Business and Psychology, 24, 373-386.

Robie, C., Brown, D. J., & Beaty, J. C. (2007). Do people fake on personality inventories? A verbal protocol analysis. Journal of Business and Psychology, 21, 489-509.

Salgado, J. F. (2003). Predicting job performance using ffm and non-ffm personality measures. Journal of Occupational and Organizational Psychology, 76, 323-346.

Schmidt, F. L., Shaffer, J. A., & Oh, I.-S. (2008). Increased accuracy for range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. Personnel Psychology, 61, 827-868.

Sjöberg, L. (1980). Similarity and correlation. In E.-D. Lantermann & H. Feger (Eds.), Similarity and choice (pp. 70-87). Bern: Huber.

Sjöberg, L. (1982). Attitude-behavior correlation, social desirability and perceived diagnostic value. British Journal of Social Psychology, 21, 283-292.

Sjöberg, L. (2008). Bortom big five: Konstruktion och validering av ett personlighetstest. (beyond big five: Construction and validation of a personality test). (SSE/EFI Working Paper Series in Business Administration No. 2008:7). Stockholm: Stockholm School of Economics.

Sjöberg, L. (2010/2012). A third generation personality test. (SSE/EFI Working Paper Series in Business Administration No. 2010:3). Stockholm: Stockholm School of Economics.

Sjöberg, L. (2012). Skönmålning i personlighetstest. (faking in personality tests). Sse/efi working paper series in business administration, no. 2012:2

Sjöberg, L. (2013). Understanding personal potential: The upp test. Manual. Version 1.0 Stockholm: Psykologisk Metod AB.

Sjöberg, L. (2015). Correction for faking in self-report personality tests. Scandinavian Journal of Psychology, 56, 582-591.

Sjöberg, L., & Wolgers, G. (2012). Personlighetstestning vid antagning till officersutbildning. En jämförelse mellan cti och upp-testet i olika grader av "Skarpt läge". Ledarskapscentrum, Försvarshögskolan. Forskningsrapport F:39.

Tett, R. P., Steele, J. R., & Beauregard, R. S. (2003). Broad and narrow measures on both sides of the personality-job performance relationship. Journal of Organizational Behavior, 24, 335-356.

---

[1] In the cited meta analysis by Schmidt et al. (2008), the highest validity was for conscientiousness with a mean validity coefficient of 0.21. This is remarkably low and no better than results obtained many decades ago, long before the "Big Five" were discovered and became popular (Allport, 1921).

[2] This group consists of a quota sample of residents of Northern England.

[3] Equal weights are usually just as good as regression weights (Bobko, Roth, & Buster, 2007), and for small samples they are to be preferred.

[4] https://upptestet.se/docs/Annex_1_Assessment_Survey.doc**.**

[5] https://upptestet.se/docs/Annex_2_reliabilities.doc

[6] https://upptestet.se/docs/proxy_validities.doc