SERVICE FIRST™

# TECHNICAL MANUAL

LARRY FOGLI

ENGLISH VERSION

ASSESSIO

ServiceFirst en-US 1.0


# TECHNICAL MANUAL

## Adaptation and Validation of ServiceFirst™

Language versions included

Danish
Estonian
Finnish
Latvian
Lithuanian
Norwegian
Polish
Russian
Swedish
US English

ASSESSIO

Statistical analysis: Anders Sjöberg, Cicek Svensson
Translation: Space 360
Cover photo: © Jeff Richt, Symbal
Graphic design: Christina Aulin
Printed by: Elanders Sverige, Stockholm 2010

ISBN: 978-91-7418-360-3
Item no.: 251-201

# Content

# Introduction

This technical manual describes the validity and reliability of the international versions of ServiceFirst. The adaptation of the US version for use in other languages is described first, after which the evidence bases for reliability and validity are presented.

## The US version of ServiceFirst

ServiceFirst is a short psychometric test that measures customer service orientation or potential, and specifically those abilities that are critical in ensuring superior customer service. ServiceFirst was developed to facilitate the selection of entry-level employees for service-oriented positions (People Focus, 2006). ServiceFirst is a self-administered test that can be completed quickly and easily, as it contains forty items and takes ten minutes to complete. It measures customer service orientation by means of items that have been validated through previous research. Administration of ServiceFirst does not require any specific degrees or other academic qualification. The US version of the ServiceFirst Test Manual contains explicit administration and scoring instructions as well as answers to questions frequently asked by applicants (People Focus, 2006). ServiceFirst measures the following dimensions:

- *Active Customer Relations*. Includes seeking and acting on service/sales opportunities with customers.
- *Polite Customer Relations*. Includes demonstrating courtesy, manners, and rapport in personal interactions with customers.
- *Helpful Customer Relations*. Includes responding to customer needs by taking extraordinary actions to assist them.
- *Personal Customer Relations*. Includes showing recognition of unique customer qualities, getting to know customers by name.

The results generated from ServiceFirst include a score for each of the four sub-scales as well as an overall total score along with a hiring recommendation. Hiring decisions should be based only on the total score

of ServiceFirst, and the scale scores should only be used by the organization to identify the applicant's strengths (e.g., Active, Polite, Helpful, or Personal in their relations with customers). These sub-scale scores are intended for information purposes only.

## The Swedish version of ServiceFirst

Swedish development efforts have expanded the extensive work performed by People Focus in the US (People Focus, 2006) and presented in their ServiceFirst Technical manual. The initial assumption was that the forty items in the final English version could also be used in the Swedish one, and it was further assumed that the four dimensions selected in the American version, would also be valid for Swedish conditions and that customized versions could also be developed. All these assumptions have been examined in a series of studies intended to determine whether ServiceFirst has the qualities required for its use in Sweden. This research has been conducted in five different ways and is documented in the Swedish manual (Assessio International, 2007) and in a peer-reviewed article (Mabon, 1989). A summary of the results follows;

- The reliability of ServiceFirst was examined in two separate studies, whose results showed acceptable internal consistency.
- Three local validation studies have been performed to determine whether or not ServiceFirst has satisfactory criterion-oriented validity in a Swedish setting and their results showed evidence of criterion-oriented validity.
- The results obtained from ServiceFirst have been compared with those from other personality, interest, and ability tests, i.e., evidence based on relationships with other variables.
- Personnel economic utility calculations have shown the financial gains that can be realized by using ServiceFirst in personnel selection systems.

## The Danish, Estonian, Finnish, Latvian, Lithuanian, Norwegian, Polish, Russian, and Swedish versions of ServiceFirst

Work on developing ServiceFirst and translating it into nine new languages in Europe started in January 2006. Assessio collaborated with an energy and retailing corporation that had been using the first translated Norwegian version for some time. The positive experiences in Norway with the service recruitment process encouraged this corporation to implement ServiceFirst in the other Scandinavian countries and on their Eastern European markets.

The aim was to use the same translation agency for all language versions and to retain the same project manager there during the entire translation period. While working on the first version, efforts were made to ensure that the specific knowledge of ServiceFirst acquired by the project manager accumulated during this project. Work proceeded in parallel on the separate versions until the summer of 2006 and the same procedure was followed for all nine versions.

An English master version was produced, based on the original US version of the questionnaire, and its text was reviewed by two native English speakers and the Project Manager at Assessio. Minor changes were made to four items for the following reasons:

One of the items was written in a male mode to illustrate a neutral mode; this might have been a problem in some languages and the item was rewritten in a neutral plural mode.

- A person does not need to look out for others as long as he looks out for himself.

Was replaced by:
- People do not need to look out for others as long as they look out for themselves.

One item was grammatically slightly incorrect and “to act polite and cheerful” was changed into “to act in a polite and cheerful manner”

The next item involved using a rather untypical phrase and was changed to a more common form; “working the check stand” was changed into ”working at the checkout”.

The last item was slightly rewritten because the first part was exactly the same as an earlier item. This was done to prevent anyone thinking that the previous item had been repeated.

-You are sitting at the counter in a restaurant when the customers sitting next to you start telling you their problems. You:

On the second occurrence this was changed to:
-You are in a restaurant and the people sitting nearest to you start to tell you all about their problems. You:

After the first translation the text was checked by an external consultant who had gained qualifications in ServiceFirst with the help of Assessio, and in every case these consultants were all native speakers of the language in question. ServiceFirst was then tried out by the local HR-department at the company who beta-tested the local version, and the comments received were entered in an Excel spreadsheet file and given to the translator. The latter then communicated in English with Assessio's project manager in Norway, who either accepted or rejected these final changes and determined the final terms.

# Sources of Evidence

## Research Sample

The sample dated 2008-03-17 consisted of 38,380 individuals. All data, except for the US data reported in Table 1, were collected from Assessio International's web-based assessment platform, which can be accessed from any device with an Internet connection. All data were collected from actual personnel selection projects in nine countries (with the exception of the US data) and the bulk of it was taken from the private retail sector. The age of the sample ranges from fourteen to seventy ($M = 24.78$ $SD = 8.50$); 63% of the sample were women, who scored somewhat higher on the total score for ServiceFirst ($r = .08$; $p < .001$), and there was a small positive correlation for age ($r = .11$; $p < .001$).

# Evidence of Reliability and Validity

This section gives an account of the psychometric properties of ten versions of ServiceFirst. Reliability and validity are defined first of all, then the analysis and results are presented, and finally a summary of the findings is given.

## Definition of Reliability

Reliability is defined as the precision of test measurement and, in the case of ServiceFirst, it is based on the collection of evidence that supports the consistency of the interpretations based on the test results. A test that is completely consistent and reliable contains no errors of measurement. This situation, however, never occurs in reality, and estimating the errors of measurement in a test, makes it possible to take them into consideration in any interpretation of the results. There are various ways of estimating test reliability such as internal consistency of the items and scales. No estimate of reliability is superior to any other; different consequences and degrees of inadequate reliability have varying relevance and impact, depending on the decisions to be made on the basis of the results.

## Definition of Validity

Validity is defined as theoretical and empirical evidence that the use of a test allows correct decisions to be made. Validation of ServiceFirst is consequently based on the collection of evidence that demonstrates the correctness of the interpretations of the test results and the conclusions drawn from them. The validity of a test should be judged by the combined volume of proof and be derived from the fields in which the test is applied. Subsequent decisions that follow the application of the method may differ in kind, but individual interpretation should be supported by the documentation. There are various ways of estimating test validity such as evidence based on the internal structure, relationships with other

measurements or test-criterion relationships. No estimate of validity is superior to any other; different consequences and degrees of inadequate validity have varying relevance and impact, depending on the decisions to be made on the basis of the results.

## Evidence Based on Internal Structure and Stability of ServiceFirst

Information about the internal structure of ServiceFirst can support validity, but internal structure evidence alone is not sufficient to establish the usefulness of ServiceFirst in predicting future work performance. To investigate evidence based on internal structure for ServiceFirst, we consider:

- Internal Consistency of the total score and the sub-scales
- Test-retest for the total score
- Goodness- of-fit measures from Confirmatory Factor Analysis (CFA)
- Goodness-of- fit measures from multiple group CFA (MG)

## Internal Consistency of the Total Score

Coefficient alpha was calculated for the forty-item total score and the results showed acceptable values for the total score for ServiceFirst in all language versions (see Table 1). The lowest value was reported for the Estonian version (alpha .77) and the highest for the Polish one (alpha = .88), while the mean alpha value was estimated at .81.

## Internal Consistency of the Sub Scales

Coefficient alpha was also calculated for the ten-item sub-scale scores (see Table 1). The results showed somewhat lower coefficient alpha values ranging from .42 to .92. The only sub-scale that showed alpha values comparable to the total score on ServiceFirst was the Active sub-scale (mean value for alpha =.83). The internal consistency for Polite (.47), Helpful (.51), and Personal (.51) showed much lower values.

## Test-retest of the Total Score

The test-retest study for Sweden represents train conductors (n=27). There was an interval of 6 weeks between the administrations. The age of the sample ranged from 20 to 56 (M= 31.48; SD=11.83). Test-retest value was estimated to .85. The test-retest study for Norway represents university students (n=114). The interval between the administrations was approximately four weeks. The age for the sample ranged from 18 to 58 (M= 26.93; SD= 5.94). Test-retest value was estimated to .81.

## Summary of Internal Consistency and Stability

The results of the analysis support the recommendation in the US technical manual that *hiring decisions should be based only on the overall total test score* (People Focus, 2006). The scale scores should be used by the hiring manager solely to determine where the applicants´ greatest strengths lie (e.g., in being Active, Polite, Helpful, or Personal in their relations with customers). As mentioned above, these scale scores are intended for information purposes only.

## Goodness-of-fit measures: Confirmatory Factor Analysis (CFA)

Confirmatory factor analysis (CFA) is a factor analysis performed for the purpose of confirming a hypothesized factor structure, in this case a one-factor solution. CFA analysis was conducted with the AMOS software (Arbuckle, 2007) through maximum-likelihood estimation. The sub-scales, rather than individual items, were used as indicators in the measurement model. Sub-scales keep the requirements for sample size manageable and are more reliable than single items. They are also more normally distributed than single items, yet they still provide multiple indicators per latent construct (Bagozzi & Edwards, 1998). The raw variance–covariance matrices were used as an input for all analyses.

In table 1 the results from ten CFAs are presented. Along with the chi-square statistics six goodness-of-fit statistics are presented; GFI = Goodness-of-Fit Index; AGFI = Adjusted Goodness-of-Fit Index; CFI = Comparative Fit Index; TLI = Tucker Lewis Index; NFI = Normative Fit Index; RMSEA = Root Mean Square Error for Approximation. GFI, AGFI, CFI, TLI, and NFI values at or above .95 are generally considered indicative of good fit,

while .90 is often used as a lower limit for defining an adequate fit. For the RMSEA, values smaller than .05 indicate a good fit, and values of .08 represent an upper limit for an acceptable fit (Hu & Bentler, 1997). Although the results differ somewhat when comparing the various language versions, the overall results show an acceptable fit between the proposed one-factor model and the data. One fit index (mean TLI = .88) does not reach the mean level of .90 which is used as a lower limit for defining an adequate fit. The RMSEA value has a mean value of .12, which indicates some differences between data and the model. All other fit indices showed acceptable results.

## Goodness-of-fit measures: Multiple Group CFA

Multiple Group (MG) CFA, following the principles suggested by Cheung (2008), was performed to test for measurement equivalence across the ten versions of ServiceFirst. Measurement bias occurs when the relationship between the construct and the observed scores differs between two or more groups. For example, a test would suffer from measurement bias if observed scores reflected customer service ability for one group but not for another group (i.e., the ServiceFirst measured different constructs as a function of the specific language version). Likewise, measurement bias would also be evident if a set of scores were reliable for one group of test subjects but not for another, i.e., it measured the same customer service construct to different degrees. Measurement bias can be conceptualized as a set of questions regarding the equivalence (or invariance) of different parameters of a measurement model.

Following a baseline model that had no constraints requiring equality among the groups, constraints were imposed on configural invariance (same factor), metric invariance (same factor loadings) and invariant uniqueness (same reliabilities). Although the $\chi^2$ values were significant, the results of the fit indices can be taken as evidence that ServiceFirst is free from measurement bias across its ten different language versions. Even the most restricted model in these tests, which imposed equality constraints on the measurement error terms, still provides a close (albeit not exact) fit between the data and the model, thus suggesting that the measurement model parameters appear to be stable across language versions.

The three additional tests to assess construct-level invariance (scalar invariance, invariant factor variances and invariant factor means) show that the levels in ServiceFirst diverge to some degree across language versions with regard to intercepts, variances and means. This result suggests that, if possible, the best practice is to use local norms for the different language versions.

*Table 1. Descriptive statistics for all language versions*

| | Danish | US | Estonian | Finnish | Latvian | Lithuanian | Norwegian | Polish | Russian | Swedish | Tot. sample |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *N* | 3643 | 867 | 1354 | 292 | 1371 | 1785 | 17685 | 1391 | 817 | 9878 | 38380 * |
| **Descriptive** | | | | | | | | | | | |
| Mean (ServiceFirst) | 152.50 | 144.05 | 151.34 | 163.25 | 150.78 | 152.93 | 153.59 | 153.17 | 148.80 | 156.66 | 154.02 |
| SD (ServiceFirst) | 12.60 | 17.45 | 11.94 | 11.13 | 12.64 | 13.03 | 11.41 | 18.30 | 14.05 | 11.92 | 12.34 |
| Mean (Active) | 39.48 | 36.36 | 37.30 | 39.24 | 38.73 | 40.53 | 39.61 | 38.67 | 37.46 | 40.46 | 39.66 |
| SD (Active) | 4.65 | 6.70 | 4.59 | 4.53 | 5.06 | 4.34 | 4.15 | 6.81 | 6.10 | 4.53 | 4.59 |
| Mean (Polite) | 37.43 | 35.88 | 38.32 | 41.04 | 37.68 | 37.47 | 37.78 | 37.09 | 38.05 | 38.85 | 38.03 |
| SD (Polite) | 3.68 | 5.89 | 3.57 | 3.45 | 3.62 | 4.03 | 3.50 | 4.58 | 4.10 | 3.43 | 3.64 |
| Mean (Helpful) | 38.62 | 38.94 | 38.09 | 43.56 | 37.81 | 39.46 | 39.46 | 39.40 | 37.02 | 38.96 | 39.12 |
| SD (Helpful) | 4.20 | 5.02 | 4.54 | 3.28 | 4.11 | 4.50 | 3.98 | 5.50 | 4.73 | 4.27 | 4.26 |
| Mean (Personal) | 36.97 | 32.99 | 37.63 | 39.42 | 36.56 | 35.48 | 36.74 | 38.01 | 36.27 | 38.38 | 37.21 |
| SD (Personal) | 4.38 | 4.59 | 4.24 | 3.79 | 4.47 | 4.79 | 3.90 | 5.52 | 4.66 | 4.10 | 4.25 |
| **Internal Consistency** | | | | | | | | | | | |
| ServiceFirst | .80 | .87 | .77 | .80 | .80 | .79 | .79 | .88 | .80 | .80 | .80 |
| **Internal Consistency** | | | | | | | | | | | |
| Active | .83 | .87 | .80 | .79 | .83 | .79 | .81 | .92 | .87 | .83 | .82 |
| Polite | .44 | .66 | .45 | .48 | .40 | .48 | .43 | .48 | .45 | .42 | .43 |
| Helpful | .48 | .64 | .50 | .45 | .45 | .52 | .52 | .64 | .42 | .50 | .50 |
| Personal | .46 | .57 | .48 | .50 | .52 | .48 | .45 | .64 | .51 | .50 | .48 |
| **Confirmatory Factor Analysis** | | | | | | | | | | | |
| $\chi^2$ | 68.62 | 27.10 | 7.62 | 12.32 | 19.04 | 18.50 | 348.42 | 11.36 | 39.27 | 194.00 | 9.83 |
| *df* | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| GFI | .99 | .98 | .98 | .98 | .99 | .99 | .99 | .96 | .98 | .99 | .98 |
| AGFI | .95 | .92 | .88 | .90 | .97 | .97 | .95 | .80 | .89 | .95 | .92 |
| CFI | .98 | .97 | .91 | .95 | .98 | .99 | .97 | .95 | .93 | .97 | .96 |
| TLI | .93 | .91 | .74 | .86 | .94 | .96 | .92 | .85 | .79 | .92 | .88 |
| NFI | .98 | .97 | .92 | .95 | .98 | .99 | .97 | .95 | .93 | .97 | .96 |
| RMSEA | .10 | .12 | .16 | .13 | .08 | .07 | .10 | .20 | .15 | .10 | .12 |
| RMSEA 95% CI Lower | .08 | .08 | .13 | .07 | .04 | .04 | .09 | .17 | .11 | .09 | .09 |
| RMSEA 95% CI Upper | .12 | .16 | .19 | .21 | .11 | .10 | .11 | .23 | .19 | .11 | .15 |

* The International norm group (N = 38380) is based on a slightly different sample. The US data were exchanged with the data collected for the English version (N = 164). The reason for this was that the primarily US data were not collected for normative purposes, instead the data were used for validity reason. Therefore, the new English sample mean and standard deviation for the total score (M = 154.37; SD = 13.03) were used in the international norm group. Coefficient alpha for the total score was estimated to .76.

*Table 2.*
*Multiple Group CFA. Invariance*

| Statistics | Configural | Metric | Uniqueness | Variance | Scalar | Means |
|---|---|---|---|---|---|---|
| $\chi^2$ | 908.34*(20) | 1216.91*(47) | 3300.31*(83) | 4294.57*(92) | 9781.41*(127) | 10356.21*(128) |
| $\Delta\ \chi^2$ | — | 308.57*(27) | 2083.40*(36) | 994.26*(9) | 5486.84*(36) | 574.80*(1) |
| NFI | .97 | .96 | .89 | .85 | .67 | .65 |
| RFI | .91 | .95 | .92 | .90 | .84 | .83 |
| IFI | .97 | .96 | .89 | .85 | .67 | .65 |
| TLI | .91 | .95 | .92 | .91 | .85 | .84 |
| CFI | .97 | .96 | .89 | .86 | .67 | .65 |
| RMSEA | .03 | .03 | .03 | .03 | .04 | .04 |
| RMSEA 95% CI Lower | .03 | .02 | .03 | .03 | .04 | .04 |
| RMSEA 95% CI Upper | .04 | .03 | .03 | .04 | .05 | .05 |

* $p < .05$; df = degrees of freedom; NFI = Normative Fit Index; RFI = Relative Fit Index; IFI = Incremental Fit Index; TLI = Tucker Lewis Index; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation

## Evidence Based on Relationship with Other Variables (The Five Factor Model of Personality, FFM)

This form of evidence is based on the empirical relationship between ServiceFirst and the Five Factor Model of personality. Although ServiceFirst was developed prior to this model (Costa and McCrae, 1995), it is useful to examine empirical and theoretical links between these two models.

The first factor of the FFM, Neuroticism, measures the general tendency to experience negative reactions such as anger, guilt and disgust. In order to make the interpretation of Neuroticism easier, this factor is reversed in the present analysis, and we use the term Emotional Stability for the reversed scale. Extraverts, in addition to liking people and preferring large groups, are assertive, active, and talkative as well as cheerful, energetic, optimistic and often rather ambitious.

The third factor is Openness to experience. Open individuals are curious and like to experiment; they welcome novel ideas and unconventional values, and possess a more active imagination and aesthetic sensitivity than closed people. Agreeableness represents altruism. Agreeable individuals are typically described as sympathetic to others and eager to help them, but they also believe that others will be equally helpful in return.

Conscientiousness is conceptualized as the ability to resist impulses and temptations. The form of self-control that conscientious individuals practice is a more active process of planning, organizing, and performing tasks. People with a high degree of Conscientiousness are purposeful, strong-willed and determined, and show a great achievement drive. ServiceFirst was specifically developed to measure the abilities that are critical in delivering superior customer service, what is termed a compound variable, whose validity will be higher than any individual sub-factor when it is formed with care. To investigate the relationship between the FFM and ServiceFirst, total score data as reported by Olesen (1997) were used and the sample consisted of 124 male and 232 female undergraduate students. Table 1 (Olesen, 1997) shows the correlation between the "Five-Factor Inventory" (FFI), the short form of the US version of the NEO-PI published by Costa & McCrae in 1992, and ServiceFirst. Twelve items were used in measuring each factor of the FFM and the reliabilities (Cronbach´s alpha) of the five scales in FFM were as follows; Emotional Stability ($\alpha$ = .86), Extraversion ($\alpha$ = .76), Openness to Experience ($\alpha$ = .72), Agreeableness ($\alpha$ = .70), Conscientiousness ($\alpha$ = .84), and ServiceFirst ($\alpha$ = .87).

Table 3 shows both uncorrected and corrected correlations for attenuation. The results indicated that the main factors in ServiceFirst, according to the FFM, are Extraversion and Openness. Although these correlations show the greatest overlap between the constructs, Emotional Stability, Agreeableness, and Conscientiousness also show moderate positive relationships with the ServiceFirst total score.

## Evidence Based on Relationship with Other Variables (Meta-analysis)

This form of evidence is based on the empirical relationship of predictor scores to external variables. The strategy adopted for the local validation of ServiceFirst was concurrent validation, which involves administering the test and collecting performance appraisal data on an incumbent group of employees. Under this validation approach, the examinees have already been performing the job prior to testing. The relationship between test scores and performance appraisal ratings provides an index of the validity of the test. The customer service appraisal form was filled out by store managers to evaluate on-the-job performance (e.g., cashier/checkers and courtesy clerks).

*Table 3.*
*Relationship between FFM and ServiceFirst (N=356)*

| FFM | Uncorrected r | Corrected r |
|---|---|---|
| Emotional Stability | .22 | .25 |
| Extraversion | .41 | .58 |
| Openness | .35 | .52 |
| Agreeableness | .24 | .38 |
| Conscientiousness | .31 | .36 |

*Table 4.*
*Meta-analysis*

| Scale | k | N | Mean $_r$ | $\rho_{op}$ | $SD_{op}$ | perc $10_{op}$ | perc $90_{op}$ |
|---|---|---|---|---|---|---|---|
| ServiceFirst | 19 | 3113 | .32 | .44 | .12 | .28 | .60 |

k = number of correlations; N = Sample size; $Mean_r$ = Sample size weighted mean observed correlation
$\rho_{op}$ = operational validity not corrected for the attenuating effects of measurement error in the independent variable. *perc* $10_{op}$ 80% credibility interval (percentile 10); *perc* $90_{op}$ 80% credibility interval (percentile 90).

The summary of ServiceFirst validation studies (see Appendix B) was analyzed by following the Hunter and Schmidt (2007) meta-analysis procedures. The relationship between the ServiceFirst total score and criteria (supervisory ratings) were individually corrected for artifacts. We were able to obtain range restriction data for five of the eighteen validity coefficients included in the dataset. The variance (SD = 12.34) of the unrestricted group was taken from the international database (N = 38 380). The five studies (N = 1448) reporting standard deviations (mean SD = 13.54) exceed the standard deviation from the unrestricted group; hence we assume that there are no problems with the restriction of range in our data. This standpoint may result in an underestimation of the true validity of ServiceFirst because it is a reasonable assumption that all the companies in these studies are unlikely to hire the applicants completely at random.

As reliabilities for rater agreement for supervisory performance ratings were not available in the studies included in our dataset, it was decided to use the best available reliability estimates for supervisory performance ratings.

In fact, two large-scale meta-analyses found .52 to be the average criterion reliability estimate for supervisory performance ratings (Salgado et al., 2003; Viswesvaran, Ones, & Schmidt, 1996). Hence, it was decided to use the value of .52 as the criterion reliability estimate for all nineteen validity coefficients. As Table 1 shows, the mean observed r based on a total sample size of 3113 was .32. Correcting this coefficient for criterion unreliability increased its value to .44. Thus the population based operational validity of ServiceFirst is .44.

## The Financial Benefits of Using ServiceFirst

In the final analysis, the value of a selection procedure is not determined by its predictive validity, but by its financial benefits to the user. Utility theory, developed in the past fifty years, shows how psychometric data can be translated into economic terms (see, for instance, Cascio, 2000). An illustration of how utility theory can be applied in practice to empirical data in ServiceFirst was provided in Mabon, 1998. A general framework, which can be adapted and utilized by most test users, is as follows:

The classic Brogden – Cronbach – Glaser model states the following:

$$\Delta u = N_s \times r_{sy} \times SD_y \times \lambda/\phi - N_s \times c/\phi$$

Where $\Delta u$ is the marginal utility of a new selection procedure, $N_s$ is the number of people selected, $r_{xy}$ is the correlation between predictor and criterion (or rather the predictive improvement compared with previous procedures), $SD_y$ is the standard deviation of performance utility expressed in economic terms, $\phi$ is the selection ratio, $\lambda$ is a function of $\phi$ and c is the incremental individual cost of the new procedure. This marginal utility refers to a single year and should be increased depending on the actual length of service of the new employees.

The objective is to calculate the amount the company would stand to gain by using a new selection procedure such as ServiceFirst. Users would have access to their own empirical data for some parts of the equation, whereas for others they could rely on a rule of thumb from previous extensive studies. Hence, a company would know the number of job applicants, how many they select, and how long people tend to stay in the job. The meta-study shown above suggests that they may assume that the validity of Service First is 0.44 and they would also know the cost of the new selection procedure.

It might prove somewhat harder to determine the standard deviation of performance utility and the validity of current selection procedures. The classic assumption in the case of $SD_y$ is that it might be regarded as 0.40 x Salary, while current selection validity based on, an unstructured interview would be unlikely to exceed 0.30, which will give an incremental validity of 0.44 – 0.30 = 0.14. The better current selection techniques are, of course, the less will be the incremental value of a new test with a validity of 0.44. It is considered, however, that the relatively summary procedures currently used to select lower-level service personal suggest that few companies will have attained the relatively high validity of 0.30.

Let us take a specific example:

The entry-level service job in question pays € 20,000 per annum giving an SD of € 8,000; newcomers will tend to stay eighteen months. The company has 1200 applicants per year and they select 288. This gives a $\phi$ of 0.24 and $\lambda/\phi$ of 1.30 from the Naylor-Shine tables (cf. Cascio, 2000). The cost of testing with Service First is €10 per person, and this is taken to be the incremental cost. It could of course be maintained that the incremental cost is negative, as other expensive procedures will now be eliminated, but this figure will be retained as a safeguard.

Marginal utility can now be calculated by using the BCG formula and multiplying the first part of the equation by 1.5 to allow for period of tenure (eighteen months):

$\Delta u$ = 288 x 0.14 x 8,000 x 1.30 x 1.5 – 1,200 x 10

= 628,992 – 12,000

≈ € 617,000

Giving the above assumptions, all of which are based on direct or indirect empirical data, it appears that the company in this case would make very substantial financial gains by enhancing the validity of their selection procedure with the aid of a valid test such as Service First. The cost of testing is in fact little more than 1% of the prospective gains.

## Summary ServiceFirst

An important consideration in planning an assessment is whether sufficient validity evidence already exists to support the proposed use of a test. The availability and relevance of existing evidence and the potential information value of new evidence should be carefully weighted in designing the assessment. This manual presents existing evidence and provides support for the generalization that ServiceFirst should be used in service recruitment systems.

# References


Arbuckle, J. L. (2007). Amos 16.0 User's Guide. Chicago, IL: AMOS Development Corporation.

Assessio International (2007). Manual ServiceFirst. Swedish version. Hunter Mabon. Stockholm, Sweden: Author.

Bagozzi, R. P., & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods, 1*, 45-87.

Cascio, W. F. (2000). *Costing Human Resources. The Financial Impact of Behavior in Organizations* (4th ed.). Cincinnati, OH: South-Western College Publishing.

Cheung, G.W. (2008). Testing equivalence in the structure, means, and variances of higher-order constructs with structural equation modeling. *Organizational Research Methods, 11*, 593-613.

Costa, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment, 64*, 21-50.

Costa, P. T., Jr., & McCrae, R. R. (1992). Revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources, Inc.

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

Mabon, H. (1998). Utility aspects of personality and performance. *Human Performance, 11* (2/3), 289-304.

Olesen, E.P., McDaniel, M.A., & Snell, A. F. (1998, April). *Construct validity of customer service measures: An examination of convergent validity evidence*. Paper presented at the 13th Annual Convention of the Society for Industrial and Organizational Psychology, Dallas, TX.

People Focus (2006). ServiceFirst Technical Report: Development and Validation of ServiceFirst. Pleasant Hill, CA: Author.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., De Fruyt, F., & Rolland, J. P. (2003). A Meta-Analytic study of General Mental Ability Validity for Different Occupations in the European Community. *Journal of Applied Psychology, 88*, 1068-1081.

Schmidt, F.L., & Le, H. (2004). Software for the Hunter-Schmidt meta-analysis methods. University of Iowa, Department of Management & Organizations, Iowa City, IA 42242.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557-560.

# Appendix A

*Table A1.*
*ServiceFirst inter-correlations overall sample*

| Scale | Active | Polite | Helpful | Personal |
|---|---|---|---|---|
| Active | 1 | | | |
| Polite | .34 | 1 | | |
| Helpful | .34 | .40 | 1 | |
| Personal | .45 | .36 | .44 | 1 |

*Table A2.*
*Inter-correlations Danish version*

| Scale | Active | Polite | Helpful | Personal |
|---|---|---|---|---|
| Active | 1 | | | |
| Polite | .35 | 1 | | |
| Helpful | .37 | .43 | 1 | |
| Personal | .45 | .35 | .47 | 1 |

*Table A3.*
*Inter-correlations US English version*

| Scale | Active | Polite | Helpful | Personal |
|---|---|---|---|---|
| Active | 1 | | | |
| Polite | .33 | 1 | | |
| Helpful | .39 | .56 | 1 | |
| Personal | .44 | .41 | .53 | 1 |

*Table A4.*
*Inter-correlations Estonian version*

| Scale | Active | Polite | Helpful | Personal |
|---|---|---|---|---|
| Active | 1 | | | |
| Polite | .29 | 1 | | |
| Helpful | .17 | .41 | 1 | |
| Personal | .36 | .33 | .42 | 1 |

*Table A5.*
*Inter-correlations Finnish version*

| Scale | Active | Polite | Helpful | Personal |
|---|---|---|---|---|
| Active | 1 | | | |
| Polite | .38 | 1 | | |
| Helpful | .29 | .42 | 1 | |
| Personal | .48 | .37 | .43 | 1 |

*Table A6.*
*Inter-correlations Latvian version*

| Scale | Active | Polite | Helpful | Personal |
|---|---|---|---|---|
| Active | 1 | | | |
| Polite | .37 | 1 | | |
| Helpful | .32 | .35 | 1 | |
| Personal | .43 | .35 | .44 | 1 |

*Table A7.*
*Inter-correlations Lithuanian version*

| Scale | Active | Polite | Helpful | Personal |
|---|---|---|---|---|
| Active | 1 | | | |
| Polite | .35 | 1 | | |
| Helpful | .36 | .38 | 1 | |
| Personal | .42 | .35 | .49 | 1 |

*Table A8.*
*Inter-correlations Norwegian version*

| Scale | Active | Polite | Helpful | Personal |
|---|---|---|---|---|
| Active | 1 | | | |
| Polite | .32 | 1 | | |
| Helpful | .36 | .43 | 1 | |
| Personal | .42 | .32 | .45 | 1 |

*Table A9.*
*Inter-correlations Polish version*

| Scale | Active | Polite | Helpful | Personal |
|---|---|---|---|---|
| Active | 1 | | | |
| Polite | .42 | 1 | | |
| Helpful | .49 | .60 | 1 | |
| Personal | .65 | .54 | .61 | 1 |

*Table A10.*
*Inter-correlations Russian version*

| Scale | Active | Polite | Helpful | Personal |
|---|---|---|---|---|
| Active | 1 | | | |
| Polite | .28 | 1 | | |
| Helpful | .23 | .35 | 1 | |
| Personal | .50 | .31 | .40 | 1 |

*Table A11.*
*Inter-correlations Swedish version*

| Scale | Active | Polite | Helpful | Personal |
|---|---|---|---|---|
| Active | 1 | | | |
| Polite | .35 | 1 | | |
| Helpful | .29 | .35 | 1 | |
| Personal | .47 | .36 | .43 | 1 |

# Appendix B

*Table B1. Meta-analysis summary. Local validation studies were conducted in the USA and Sweden*

| Industry | Job | N | α | Criterion | r |
|---|---|---|---|---|---|
| Retail Supermarket | entry-level positions | 829 | .87 | overall customer service | .44 |
| Retail Supermarket | entry-level positions | 198 | .87 | overall customer service | .23 |
| Retail Nursery | entry-level positions | 114 | .87 | overall customer service | .35 |
| Telecommunications | Maintenance Administrators and System Technicians | 164 | .87 | overall customer service | .20 |
| Health Care | Nurses | 91 | .82 | overall customer service | .21 |
| | Clerical | 117 | .82 | overall customer service | .22 |
| | Manager/ Professional | 90 | .82 | overall customer service | .07 |
| | Technical | 85 | .82 | overall customer service | .10 |
| Swedish Post Office | Postal Workers | 221 | .87 | overall customer service | .30 |
| Swedish Temporary Employment Agency | Clerical | 97 | | overall customer service | .23 |
| Swedish Retail Women's Clothing | entry-level positions | 263 | | overall customer service | .18 |
| Telecommunications | Operators | 56 | .87 | overall performance | .24 |
| Temporary Agency | Clerical | 91 | .87 | overall performance | .37 |
| Retail Gas | Mini-Mart Managers | 81 | .87 | overall performance | .32 |
| Real Estate | Clerical | 131 | .87 | overall customer service | .45 |
| Banking – adapted survey | Bank Tellers | 267 | .89 | overall customer service | .36 |
| Insurance – adapted survey | Clerks | 101 | .65 | overall customer service | .31 |
| Insurance – adapted survey | Claim/service adjusters | 90 | .87 | overall customer service | .48 |
| Railroad | Train Catering Steward | 27 | .80 | overall customer service | .20 |