Balance Impossible? Between Customizability and Comparability of Employee Satisfaction Surveys

Christoph HARTEBRODT[*] – Yvonne CHTIOUI

Forest Research Institute Baden-Württemberg (Germany)

Abstract – In a significant share of cases, when multi-objective management systems (e.g. Balanced Scorecard, Quantum Performance Measurement) are in use, the goal of employee satisfaction is part of the set of strategic goals. Quite frequently, this goal is flanked by an indicator, mostly a so-called employee satisfaction index, which is frequently derived from an employee satisfaction survey (ESS). On one hand such a survey has to reflect the characteristics and structure of the enterprises, on the other hand it is known that the question for intercompany-comparison turns up immediately after the disclosure of the results. The paper discusses the results of a meta-analysis of ESS in seven of the larger forest enterprises in Germany. The main topics are underlined and the problems of comparability of customized ESS are shown, while focusing on both wording and scales used for their measurement. A methodological approach of dealing with various scales is discussed based on the results of inter- and intra-company ESS. A vision of a common ESS framework is outlined.

intercompany comparisons of employee satisfaction / transformation of Likert-scales

Kivonat – Lehetetlen középút? A munkahelyi elégedettségi felmérések testreszabhatósága és összehasonlíthatósága. Az esetek jelent s részében, amikor többcélú vállalatirányítási rendszerr l (pl. kiegyensúlyozott stratégiai mutatószám-rendszer, quantum teljesítmény értékelés) van szó, a stratégiai célrendszer része a munkahelyi elégedettség. Gyakran, ezt a célt az un. munkahelyi elégedettségi indexszel jellemzik, amelyet a munkahelyi elégedettségi felmérésb l származtatnak. Egyrészt egy ilyen felmérés vissza kell, hogy tükrözze a vállalkozás egyedi szerkezetét és jellegzetességeit, másrészt tudható, hogy a vállalkozások közötti összehasonlíthatóság igénye azonnal megjelenik, amint a vizsgálat eredményei megszületnek. Ez a cikk a munkahelyi elégedettségi felmérések meta-elemzését tárgyalja Németország hét nagyobb méret erdészeti vállalata példáján. Az írás bemutatja a f bb témaköröket és az egyedi munkahelyi elégedettség felmérések össze-hasonlíthatóságát az alkalmazott fogalmak és a mérési skálák tekintetében. A különböz mérési skálák összevetésével kapcsolatos módszertani kérdésekre a vállalkozáson belüli és vállalkozások közötti munkahelyi elégedettség felmérések alapján szintén kitérnek a szerz k. A munkahelyi elégedettség felmérések jöv beli közös általános kereteit felvázolják.

munkahelyi elégedettség felmérések vállalkozások közötti összehasonlítása / Likert skála transzformációja

---

[*] Corresponding author: christoph.hartebrodt@forst.bwl.de; Wonnhaldestraße 4, D-79100 FREIBURG

1 INTRODUCTION

1.1 Preface

Employee satisfaction is an objective that is on the agenda in almost all of the larger institutions. Especially if there are multidimensional (performance) management systems in use, it can be seen that the goal of 'employee satisfaction' is typically part of the set of strategic targets. A higher share of these 'modern' management systems, e.g. (sustainability) balanced scorecard, use indicators for the measurement of successes or failures in executing the strategy. In the case of the employee satisfaction, this is mostly an index value and in almost all cases the methodological approach is an employee satisfaction survey (ESS). The literature highlights, on one hand, the need for the characteristics of the individual organisation to be addressed in ESS; on the other hand, the matter of comparability is discussed. As participation of targeted groups is seen as a precondition for the later acceptance and success of the following measures, the issue of comparability is generally less important in the period of the design of the ESS. This changes notably when the results are published. After a short discussion of its own results, unavoidably the question of: "How does our company compare to other institutions?" arises. The question and core topic of the paper is whether a balance between customizing and comparability is possible.

1.2 Contents

Intensive research has been undertaken in past years in relation to the topic of employee satisfaction surveys (ESS). There is a consensus that these surveys cover on average all topics, which are evaluated as being important for employee satisfaction (Borg 2003). However, the opinions regarding what the minimum requirements of such a survey should be differ strongly. Borg (2003) underlines that most surveys include a combination of standard topics and questions related to the particular situation of the institution. This is even the case when the motive for the survey seems to be very specific. This can be explained by the fact that a narrow range of topics can result in a more or less strong refusal by the members of the institution. In addition, it can be stated that there is no unique recommendation on what the set of standard topics should be. Nonetheless, desktop research showed that there is a discussion about some type of an intersection-set of questions, which appears in a relevant number of questionnaires (inter alia. Balmer et al. 2000).

Borg (2003) classifies 11 standard contents from the point of view of the members of the institution and 9 performance and strategy-related contents, which are of particular relevance for the institution as such. He derives a structure for three areas of interest from these findings. Hossiep & Frieg (2008) undertook a study of 820 of the largest enterprises in Germany, Austria and Switzerland. They identify 19 standard and performance topics and 9 psychological issues, which are addressed in ESS. If this set of topics is split up into a stratum, which occurs in more than 50% of the surveys and one which is less frequent, a three-part classification emerges. This three-partite structure is basically confirmed by Bösch (2011), who identifies three standard topics that are part of almost every ESS, five that are frequent, and seven contents that are related to psychological, social or ethical issues. Fischer et al. (2008) do not define areas of interest. They define 12 fields of questions, which are seen as being the most important key drivers for satisfaction and commitment of employees. Domsch & Ladwig (2000) and the European Foundation for Quality Management (EFQM2013) define 9 areas of interest; from which five (so-called enablers) have a stronger relationship to the issue of employee satisfaction.

From the findings above, it can be subsumed that there is some type of an understanding of the important aspects, which are part of ESS, but the opinions about how to structure these

topics differ notably. As a reason for this heterogeneity, two main aspects seem to be of special significance: the need to develop a case-related solution (a) and the participation of the employees (b).

(a) Evaluation theory tells that core objective of every evaluation is answering the specific questions (Rossi et al. 2004). Borg (2003) and Fischer et al (2008) point out that the content of any kind of ESS should be related to the objective of the study, which means that the mere use of a standard set of questions would be misleading.

(b) As participation of employees in the stage of the development of ESS is recommended in order to cover all relevant aspects from the point of view of the employees, some very specific contents and / or questions are to be expected. Unless this is seen as a constraint with the later comparability of the results, this participation is seen as a precondition for the acceptance of the results and the consecutive measures (ibid.).

1.3 Scales

As shown for the contents, there is also no consensus about the use of scales in ESS. The spectrum ranges from the statement that there are no good or bad scales, but appropriate and inappropriate scales related to the respective questions (HBS 2002), to specific findings on what the best solution is (e.g. Cummins and Gullone (2000): no five point (odd) scale; e.g. Sturgis et al. (2014): pro neutral alternative). There is a certain consensus that five-point Likert scales are frequently used, and one should not exceed nine categories (Eurostat 2004). Less homogeneity can be found related to the question of whether balanced or unbalanced scales, or even or odd scales were favourable and also about the meaningfulness of a 'don't know' alternative.

1.4 Calculation of Index Values

It can be shown that there are different types of index values which are subsumed under the headline of ESS. Bösch (2011) differentiates between four types: (1) satisfaction-index, (2) leadership-index, (3) commitment-index and (4) acceptance-index. Evidence shows that there are various sets of questions, which are related to the respective indices. However, there is a wide overlap between sets related to different indices and evidently no consensus exists in terms of which question supports which type of index-value. Despite the fact that there is a discussion about the accuracy of deriving arithmetic means from ordinal and/or nominal scale by coding them with discrete numeric values, this is applied in almost all cases. After coding the questions, two basic approaches of deriving an index value from the questionnaires are used. The first one can be titled as the 'all questions approach'. Here the results of all individual questions are included in the calculation of an index, partially after calculating means of sections and combining these means to an overall index. The second approach can be subsumed under the 'index-question' method. Here a special set of questions, which are used to derive the overall satisfaction index, is included. These questions are frequently intended and used to include the antecedent section too.

2   MATERIAL AND METHODS

2.1 Material

As the study was embedded in an ESS-project under the umbrella of a Sustainability Balanced Scorecard (SBSC) project, which is a management approach more suited mainly for medium-sized and larger forest entities, it was decided to include only larger forest organisation in the

German-speaking regions. It was possible to identify six ESS studies from the past six years, which could be compared with the one of the state-owned Forest Enterprise of Baden-Württemberg from 2013. The comparative studies are listed in *Table 1*.

Table 1.   List of comparative studies

Institution	Year
State Forest Enterprise Saxony-Anhalt	2007
State Forest Enterprise of Hessen (HessenForst)	2009
State Forest Enterprise of Bavaria (BaySF)	2010
Forest Research Institute Baden-Württemberg (FVA)	2010
State Forest Enterprise Mecklenburg-West Pomerania (Landesforsten M-V)	2010
Forest Research Institute of North-Western Germany (NW-FVA)	2011

2.2 Methods

The study focusses mainly on four topics. It was of special interest whether:
- the ESS forest case study addresses topics that are seen as being important in the general literature about ESS
- the wording of questions allows for a comparison between individual questions
- the scales applied allow for a direct comparison
- a normalisation of scale can be used in the case that comparable questions are evaluated with different scales.

The methodology applied in respect to these areas of interest is described below.

2.2.1   Topics

Different authors used a "frequency-approach" to classify the importance of the topics; this method was applied in this paper in order to identify the most important topics for further research. Based on desk research, six classifications were combined in a meta-analysis. Borg (2003), Hossiep & Frieg (2008) and Bösch (2011) use a three-partite scale, which was included directly into the meta-analysis with three, two, and one points respectively. Domsch & Ladwig (2000) differentiate their results into core topics and sub-items. Core topics were valued with two points, whereas sub-items receive a one point value. Bröckermann & Müller-Vorbrüggen (2010) and Fischer et al. (2008) define a limited set of key drivers which were coded with two points. The standards of the EFQM were – due to their international significance -ranked higher. In the case that a topic is listed in the standards of EFQM, three points applied. In a second step, the comparative questionnaires were analysed to determine if they contained questions related to the single topics that are included in the questionnaire; if they did, those questions were counted. The sum of points related to the individual topics was used to derive a ranking of important topics. For the most important topics, the number of questions in the comparative studies was analysed.

2.2.2   Questions and their Wording

The 'handbook for […] questionnaire-development' issued by Eurostat (2014) states: "Minor changes in wording can also have a significant effect on responses". Together with findings that standardized questions (questionnaires) are preferable whenever comparability is needed, it can be stated that it is widely accepted that the results of ESS can only be used for comparison studies in cases where the wording of the individual questions is more or less identical. Therefore, the questions in the ESS were assigned to the list of topics. Using text analysis, it was tested to see if a direct comparability of the wording was feasible or not.

2.2.3 Scales
The scales used in the questionnaires were analysed. Of special interest were:
- number of categories
- type of scale (ordinal / cardinal)
- use of a "don't know" category
- even or odd number of categories
- the use of a neutral category
- whether the wording for the scales applied was balanced or unbalanced

2.2.4 Normalisation
Based on the analysis of the questionnaires, a small set of questions from two comparative studies and the ESS from ForstBW from 2013 could be identified; these showed a sufficient similarity of wording, but different scales. For this small subset, the results have been normalised using formula (1):

$$\frac{x_i - x_{min}}{x_{max} - x_{min}} * 100 \tag{1}$$

x_i = Likert code resp. mean of Likert code
x_{max} = Maximum Likert code
x_{min} = Minimum Likert code

The factors derived from this normalisation were applied to the results of the individual questions in the ESS. *Table 2* gives an example of this kind of normalisation.

Table 2. Normalisation of different scales

Likert code	% Value	Likert code	% Value
6	100	5	100
5	80	4	75
4	60	3	50
3	40	2	25
2	20	1	0
1	0		

3   RESULTS

3.1 Topics

In total, 44 topics were identified from which 10 reached more than half of the maximum sum of points derived from the six studies included, and were furthermore used as the most important topics. A second group, which reached a quarter to half of the maximum value, is depicted as well (*Table 3*). Here we found 16 topics. At the bottom of *Table 3* the topics of lower importance (three points and less) are listed.

Table 3. Priority of topics in ESS studies

Topics	EFQM (from Bösch 2011)	Bösch	Bröckermann & Müller-Vorbrüggen	Domsch & Ladwig	Borg	Hossiep & Frieg	Max 16
Leadership / immediate superior	3	3	2	2	3	3	**16**
Cooperation / team (colleagues, service providers dept.)	3	1	2	2	3	3	**14**
Working conditions and safety	3	1	2	2	3	3	**14**
Advanced training / prospects		3		2	3	3	**11**
Functions and duties / objectives		1		2	3	3	**9**
Pay / statutory benefits / gratuities		1		2	3	3	**9**
Staff retention / fluctuation	3	1	2			3	**9**
Working atmosphere	3	1	2		3		**9**
Communication / information		1		2	3	3	**9**
Processes / internal organisation	3	1	2		3		**9**
Customer orientation / customer retention		1	2		2	2	**7**
Job satisfaction / overall satisfaction		1	2	1	3		**7**
Identification / commitment / emotional bonding		1	2		3		**6**
Assertion of strategy		1			2	3	**6**
Image / communication /attractiveness		2		1		3	**6**
Management of innovations		1			2	2	**5**
Motivation, performance / productivity				1	2	2	**5**
Workload/ stress / burnout		2		1	1	1	**5**
Equitableness and acceptance		1		1	1	1	**4**
Empowerment/ accountability and freedom / participation		2			1	1	**4**
Self-respect and perspectives		2		1	1		**4**
Quality (products/ services)					2	2	**4**
Motivation		1	2	1			**4**
Work-life balance		2			1	1	**4**
Change management					2	2	**4**
Project management					2	2	**4**

Topics of lesser importance

Uncertainty of employment; Deficits / Potentials for improvement;
Change of legal status, Reorganisation, Fusion; Engagement; Discrimination;
Confidence in management und enterprise; Working hours schemes; Media; Social capital;
Diversity; Mobbing; Demographic change; Confidence; Management instruments;
Health management; Career planning / Career supervision; Management trends;
Psychological contracts

The analysis of the number of questions related to the most important topics used in the ESS showed that seven out of ten of the very important topics are addressed in the forest institutes' ESS as well. However, it has to be kept in mind that the intensity, in terms of number related questions, differs notably. The topics "payment, statutory benefits, gratuities', 'staff retention, fluctuation' and 'working atmosphere' are covered weakly in the studies (*Table 4*).

Table 4   Coverage of key topics in ESS in forest institutions

	Bavaria 2010	Hessen 2009	Saxony-Anhalt	Mecklenburg-West Pomerania	North-Western Research I.	Research I. Baden-Württemberg	Average Number of Questions
	Numbers of Questions						
Leadership / line manager	16	11	4	9	4	5	8.2
Cooperation / team (colleagues, service providers, departments)	9	8	1	0	1	3	3.7
Working conditions and safety	5	3	1	11	4	7	5.2
Advanced training / prospects	4	5	5	6	2	3	4.2
Function and duties / objectives	5	6	3	8	1	5	4.7
Pay / statutory benefits / gratuities	1	2	1	0	0	0	0.7
Staff retention / fluctuation	0	0	1	0	1	0	0.3
Working atmosphere	3	1	1	0	2	2	1.5
Communication / information	16	9	3	11	7	4	8.3
Processes / internal organisation	8	4	5	3	8	5	5.5

3.2 Wording

In the following table, the translation of the set of questions used for two topics is shown. It appears that the questions can differ widely and that only in a small share a direct comparability can be undertaken (*Tables 5 and 6*).

Table5.   Examples for the different wording of questions related to the topic 'Leadership / Direct superior'; Subtopic appreciation

Original wording in German	Equivalent wording in English
Das Verhalten meiner direkten Führungskraft mir gegenüber ist von Wertschätzung geprägt.	The behaviour of my direct superior is based on respect.
Mein Vorgesetzter spricht seine Anerkennung aus, wenn ich gute Arbeit geleistet habe.	My direct superior expresses his/her appreciation, when I have done a good job.
Ich bin mit der Wertschätzung meiner Arbeit durch meinen direkten Fachvorgesetzten zufrieden.	I am satisfied with the appreciation shown by my direct superior.
Ich bekomme Lob und Anerkennung von meinem direkten Fachvorgesetzten gezeigt.	I receive praise and appreciation from my direct superior.
Werden von ihrer Führungskraft gute Leistungen erkennbar gewürdigt?	Does your manager clearly display appreciation for a good performance?
Mein Vorgesetzter lässt mich auf verschiedene Art wissen, dass meine Leistungen anerkannt werden.	My superior lets me know through different ways that my good performance is recognised.
Erkennt ihr Vorgesetzter gute Leistungen lobend an?	Does your superior recognise a good performance?

*Table 6.   Example for different wording of questions related to the topic
             'Cooperation / Team (colleagues, service providers, departments)'*

Original wording in German	Equivalent wording in English
Zusammenarbeit zwischen den Organisationseinheiten.	Cooperation between the individual units.
Die Zusammenarbeit meines Teams mit anderen Arbeitsbereichen innerhalb der Dienststelle ist so, dass wir gute Arbeitsqualität erbringen können.	The cooperation between my team and other units works in a way that we can achieve a good working quality.
Mit anderen Fachbereichen wird gut zusammengearbeitet	The cooperation with other units works well.
Die Arbeit in meinem Team funktioniert gut	My team works well together.
Wie zufrieden sind sie insgesamt mit der Zusammenarbeit mit Kollegen aus andern Abteilungen/Teams?	How satisfied are you in general with the cooperation with colleagues from other departments?
Zwischen meiner Abteilung/ Arbeitsgruppe und anderen Abteilungen/ Arbeitsgruppen besteht eine gute Kooperation.	There is good cooperation within my department / group and other departments / groups.
Wie beurteilen sie die Zusammenarbeit mit den Kollegen anderer Abteilungen/ Gruppen?	How do you assess the cooperation with colleagues from other departments / groups?
Zusammenarbeit mit anderen Abteilungen/ Teams (internen Kunden oder Lieferanten)	Cooperation with other departments / teams (internal customers or providers)
Aus meiner Sicht ist die Zusammenarbeit a) innerhalb meiner Abteilung reibungslos b) zwischen den Abteilungen gut.	From my point of view the cooperation a) in my department works unobstructed b) between the departments is good?
Die Zusammenarbeit zwischen der Zentrale und den Forstbetrieben funktioniert reibungslos.	The cooperation between the general office and the local units runs smoothly.
Internationale Zusammenarbeit: a) Ist die Zusammenarbeit mit Kollegen in anderen Ländern ein Bestandteil ihrer täglichen Arbeit? b) Wenn ja, funktioniert diese internationale Zusammenarbeit gut?	International cooperation: a) Is cooperation with colleagues in other countries a part of your daily work? b) If yes, does this cooperation work well?

The six questionnaires had 521 subtopics in total and were grouped in the 44 topics listed above. For 76 subtopics, more than one type of wording could be identified. The number of different wordings was 11 maximum; however, two to three wording alternatives for one subtopic cover more than 60% (*Table 7*).

Due to the qualitative nature of text-analysis, the number and share of comparable questions, which is also shown, allows for only a rough estimate about the possibilities of direct comparability of questions. The number of wordings, which allow intercompany benchmarking, normally does not exceed 3; in roughly two-thirds only one opportunity for subtopic-related comparison exists. Recalling that additionally a use of different scales applies (see also *Table 8*), the number of cases in which the wording and scale allow for direct comparison tends to be insignificant.

Table 7. Number and shares of differently worded questions related to a subtopic

Number of Wordings	11	8	7	6	5	4	3	2	Sum
Absolute Frequency [N]	2	2	3	6	9	7	14	33	76
Share [%]	2.6	2.6	3.9	7.9	11.8	9.2	18.4	43.4	100
Thereof number of subtopics with direct comparable wordings									
5			1	1					2
4				1					1
3		1	1	2	2	4	1	1	12
2				3	4	3	4	11	25
Share of wordings, basically suited for intercomp. benchmarking	14.3	50.0	52.4	44.4	44.4	32.1	26.2	33.3	36.4

3.3 Scales

Table 8 shows the number of different types of scale which were used in the six questionnaires.

Table 8: Types of scales

	Alternatives						Ordinal Scale	Nominal Scale	Symmetry		Neutral Alternative		"Don't know" Alternative	
	2	3	4	5	6	>6			yes	no	yes	no	yes	no
1				x				x		x		x		x
2		x						x		x		x		x
3	x							x		x		x		x
4	x							x	x			x		x
5	x							x		x		x		x
6		x						x		x		x		x
7			x					x		x		x		x
8				x				x		x		x		x
9					x			x		x		x		x
10						x		x		x		x		x
11				x			x		x			x		x
12				x			x			x	x			x
13			x				x		x		x			x
14			x				x		x		x		x	
15			x				x			x		x		x
16		x					x		x			x	x	
17		x					x			x		x		x
18		x					x		x			x		x
19		x					x			x	x			x
20		x					x			x		x		x
21		x					x			x		x	x	
22		x					x		x		x			x
23	x						x		x			x		x
Sum	4	5	5	5	3	1	13	10	8	15	5	18	3	20

The analysis of the applied scales shows tremendous variability. Even in the limited number of questionnaires we found 23 types of scale, without mentioning these questions which are related to demographic aspects. The number of alternatives shows an almost equal distribution between 2 and 5; 6 alternatives was less frequent and more than 6 alternatives was applied only one time. Roughly 60% of the scales are ordinal scales. From the 13 ordinal scales, 7 are balanced in the sense that the wording is identical except for the word that indicates whether the answer is proven true or false (e.g. totally agree / widely agree / widely disagree / totally disagree). Most parts of the ordinal scale were so-called 'forced scales', where no neutral position is offered and where the respondent has to make a choice whether he agrees or not. Only 3 types of scales offered the "don't know" alternative.

3.4 Comparability and Benchmarking

The results presented above gave hints that the level of comparability could be low when the search for benchmarks for the 2013 ForstBW ESS was started. This expectation proved true. It was possible to identify 12 questions, which could be used for intra-company benchmarking using the results from 2001 and 2013. A sufficient comparability that could be used for comparison with the two other forest institutions was observed in only 9 questions. As we warranted anonymity, these institutions are characterised with 'Org. 1' and 'Org.2'. In both cases the scale differed and a normalisation was carried out as described above (see *Section 2.2.4*).

The *Figures 1 and 2* depict the results of the inter- and intra-company comparison.
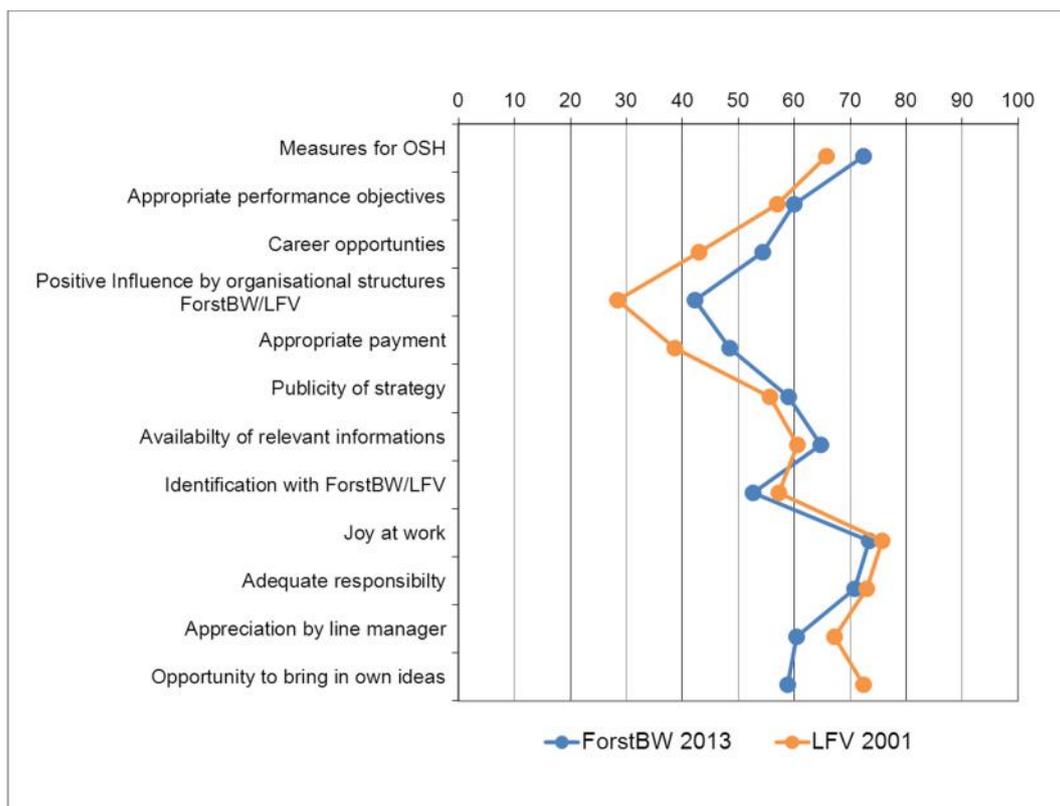

Figure 1. Intercompany benchmarking

It is clear that the patterns of the answers given by the employees of ForstBW in 2001 and 2013 are quite similar. In the year 2001 a five-point Likert scale was used, whereas in 2013 a six-point scale was applied. After the normalisation to a '100% scale', we found

values that are better and some that are worse. This can be seen as a strong indication that there is no bias caused by the normalisation. The absolute difference does not exceed 14%. The averages are very similar (ForstBW 2013 = 60, LFV 2001 = 58), that is, the coefficient of variation is smaller to some extent in 2013 (0.15) if compared with 2001 (0.24).

The intercompany differences are shown in *Figure 2*. The differences of the overall means are insignificant if comparing ForstBW with Org. 1 (1%), but can reach a notable amount in comparisons with Org. 2 (7%). The absolute values show the same characteristic. There are similarities and differences. With 27% maximum, the latter one reaches a higher level compared to the values of the intra company comparison. The coefficients of variations are at an astoundingly equal level (0.12, 0.12, 0.11). Again, it can be concluded that the normalisation allows for similar findings as well as the detection of various levels of satisfaction.
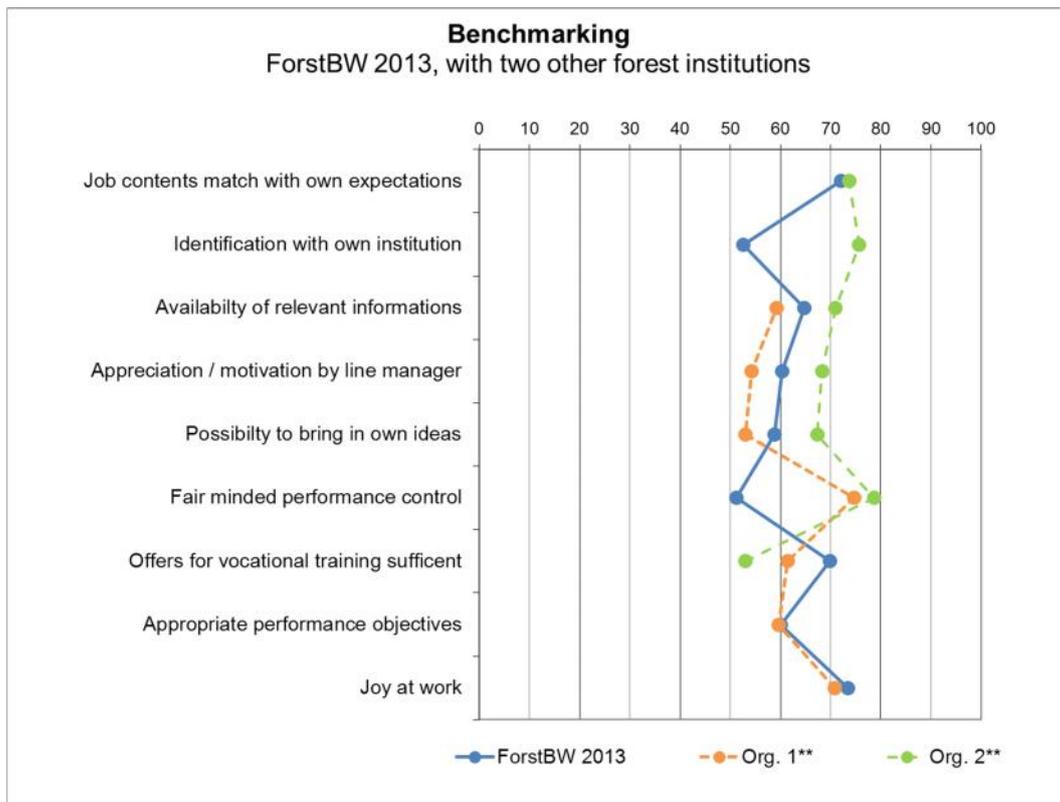

Figure 2. Intercompany comparison

4 DISCUSSION

4.1 Topics

The choice of topics within the individual ESS already shows a wide variability. Even from the 10 most important topics only 6 were addressed with at least one question in all surveys (*Table 4*). An analysis of the topics of middle and lower importance (not depicted in the paper) shows that a case-related choice of the contents prevails. Thus, it can be stated that forest ESS show the same selective or institution-related approach as described for ESS in other sectors. As participation of the employees during process of development of an ESS plays a relevant role in almost all cases, this situation is not surprising. As shown above, the number of questions related to the different topics varied tremendously. Consecutively, the importance of the topics can be very different in all cases where the satisfaction index is

calculated on the basis of the individual questions. The selection of topics and questions included in the indices is related to the total number of questions in the individual ESS, which varies too. Therefore, it has to be stated that the comparison of the overall indices should be avoided due to the fact that the issues addressed and the intensity of the ESS cannot be compared at a satisfactory level.

4.2 Wording

The analysis of the wording in the questionnaires reveals that there is no use of any kind of standards and even no visible desire to use wording from other ESS. It can be assumed that the design of the questionnaires is driven more by the personal experience and opinions of the developers, be it that this has been developed and carried out by the institution itself or from a consultant involved. In the latter case, it is rather probable that the standards of the consultant influence the wording, because this reduces the time and effort needed to develop the questionnaire. A tendency that the consultants prefer their own wordings would be supported as well by the fact that the offer of benchmarking data can be a profitable side business after the ESS. With regard to the forest ESS, it has to be stated that there were almost no identical questions and even the share of comparable questions is low. Together with a finding that the wording does not always meet technical standards, such as addressing only one topic in a single question or using only positive or negative statements in a section of a questionnaire, it has to be concluded that the chance for comparisons on the level of individual questions is limited as well.

4.3 Scales

The results of the scales applied in the six ESS used for a comparison to the ESS of ForstBW carried out in 2013 were almost self-explaining as 23 different scales could be identified in the underlying questionnaires. It has to first be questioned whether the questionnaires were designed professionally. The number of scales and changes of scale should be limited, because there is always the risk that the respondent does not realize this change and gives his answers on the basis of the scale of the previous sections. As there is almost no chance to verify whether the scales have been used properly, the results can be notably biased.

4.4 Comparability and Benchmarking

The possibilities for an in-depth benchmarking are very limited. This applies especially for the benchmarking of the employee satisfaction indices. The selection of topics and number of questions varies widely and even the way in which the indices were calculated can differ. Together with the use of different scales, partially more than one scale is used in a single ESS questionnaire; rendering it impossible to make assumptions about the differences of the overall satisfaction in the different forest institutions. This leads to a situation that benchmarking data are more or less unavailable, or have to be bought from consulting enterprises unless it is not known whether these data are comparable or not. It can be assumed that the price of the benchmarking data notably increases the costs of the whole analysis.

   The analysis of the scales offered a detailed insight into the heterogeneity of the individual questionnaires and the alternatives for responses. Recalling that only seven questionnaires are included in the analysis, it became obvious that in most questionnaires a multitude of scales is applied. An in-depth analysis of the questionnaires shows that there is frequently a change of scale from one part of the questionnaire to the next. The scales are changed partially even from question to question. This conflicts with the dominant recommendation in the literature that the change of scale has to be avoided to a maximum extent. However, there are hints that this problem is limited in questionnaires that are used to

develop some kind of employee satisfaction index. Here the developers were more aware that a change of scale can provoke errors during the completion of the questionnaire. Nevertheless, it has to be stated that the degree of professionalism could be notably improved. The wish to evaluate the opinion in respect to very specific topics is not necessarily linked to the need to use different answer-scales.

However, there is one circumstance that helps to carry out comparison of equally worded questions in the case that only the number of categories applied differs. Recalling that the number of questions included in an inter- and intra-company comparison was very limited (9 respectively 12questions, see above), there are strong hints that a normalisation of the scales could be applied. The ESS in ForstBW (2013) and LFV (2001) was executed in almost the same situation about three years after a harsh reorganisation and with the awareness that the degree of dissatisfaction was high. As more than two-thirds of the people of the study overlapped, the basic hypotheses must be that no larger differences in the overall satisfaction are given, but variations in respect to different items are to be expected. The results achieved above, after the normalisation of a five- and six-point Likert scale, have basically proven these hypotheses.

4.5 Summary

The results reveal that forest ESSs are highly customized at the moment. The possibilities to make a meaningful benchmarking are limited to single questions and even here, some, but no significant, uncertainties remain in cases where different scales had been applied. This situation is mainly driven by the requirement to include the target groups in the development process of the ESS. In addition, it cannot be excluded that the consultants force this too, because any use of standardized questionnaires would reduce the need to involve them. Moreover, the sale of benchmarking results is, at least potentially, a side-business too; this furthermore reduces the interest of consultants to use standard questions or index calculations as well. Thus, it must be stated that the value of the ESS is at present limited to an intra-company time series. However, it has to be kept in mind that organisational changes are permanent concomitants that may hinder or prevent the realisation of meaningful replication of the ESS.

The results reveal that there are topics that are of interest in most forest enterprises and that the present wording and design of the questionnaires does not always reflect the best practice standards. Consecutively, it can be assumed that the choice of professionally-worded questions from a publicly available set of questions could improve the individual ESS without necessarily hampering the requirement to cover the issues that are of special interest in the respective institution.

5   CONCLUSION AND OUTLOOK

At present it can be stated that in the German-speaking region we found customized ESS with very limited possibilities to undertake any kind of benchmarking. Furthermore, a large potential of these surveys, which are always cost-intensive undertakings, is not used. At least two approaches that can lead to a stepwise improvement in possibilities for intercompany comparability can be identified.

A kind of database, in which properly worded questions related to individual topics are available during the design of new ESS, can facilitate and speed-up the process of the development in a way that the time consuming wording of questions is replaced by a choice of questions. In the case that at least the former users of these questions, or even better (anonymized) benchmarking data are available in the database too, a cost-free and reliable

comparison becomes basically achievable. Identical scales are useful; however, the normalization of scales seems to be a feasible approach to bridge that gap.

In the case that there are already existing ESS in an institution, any change of the structure of the questionnaire or even questions and their wording must be discussed carefully, because an intra-company time series of identical repetitive ESS is a value as such. Here, an inclusion of a set of additional index questions related to the most important topics (see *Table3*),which can be used to derive a sort of intercompany standard for employee satisfaction indices, could offer the chance to compare at least the overall employee satisfaction or the satisfaction in some thematic areas. These index questions could be seen as a subset out of the database discussed above.

Such a subset could be used as well for intermediate "snapshot-ESS" between the more comprehensive and therefore more expensive regular ESS, which are frequently carried out in longer time intervals.

The balance between customizing and benchmarking is not given yet. However, it seems possible to make progress using a more incremental approach without risking the loss of the own data and experiences. Therefore, it is less of a technical problem to come closer to a balance; rather, desire is needed to enhance the transparency by using benchmarking data.

REFERENCES

BALMER, R. – INVERSINI, S. – PLANTA, V. A. – SEMMER, N. (2000): Innovation in Unternehmen: Leitfaden für Selbstbewertung in KMU. Hochschulverlag an der ETH. Mensch, Organisation und Technik, Bd. 26. 198 p, with extensive set of appendices.

BÖSCH, W. (2011): Praxishandbuch Mitarbeiterbefragungen. Praxium-Verlag, Zürich, 295 S.

BORG, I. (2003): Führungsinstrument Mitarbeiterbefragung. Theorien, Tools und Praxiserfahrungen. 3. überarb. und erw. Auflage, Hogrefe-Verlag, Göttingen, 429 S.

BRÖCKERMANN, R. – MÜLLER-VORBRÜGGEN, M. (2010): Handbuch Personalentwicklung. Die Praxis der Personalbildung, Personalförderung und Arbeitsstrukturierung. Schäffer-Poeschel Verlag Stuttgart, 798 p.

CUMMINS, R.A. – GULLONE, E. (2000).Why we should not use 5-point Likert scales: The case for subjective quality of life measurement. Proceedings, Second International Conference on Quality of Life in Cities (pp.74-93).Singapore: National University of Singapore.

DOMSCH, M. E. – LADWIG, D. H. (2000): Handbuch Mitarbeiterbefragung. Springer Verlag, Berlin, 435 S.

EFQM (2013): EFQM Excellence Modell 2013, EFQM, Brüssel. 58 pp.

EUROSTAT (2004): (complete list of authors available from the author) Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System (2004 Edition), 142. p. and additional management summary http://ec.europa.eu/eurostat/ramon/statmanuals/files/Handbook_of_Practices_for_Quest.pdf, (accessed 18.09.2014).

FISCHER, J. – STAMS, A. – TITZKUS, T. (2008): Mitarbeiterzufriedenheitsanalyse. In Bröckermann, R. (2008): Handbuch Personalentwicklung. Die Praxis der Personalbildung, Personalförderung und Arbeitsstrukturierung. 2. Auflage. Schäffer-Poeschel Verlag, Stuttgart, 721 S. S. 305–321

HBS (Hans Böckler Stiftung((2002): Mitarbeiterzufriedenheit – Arbeitspapier 54 des Fachausschusses 2/99. Eigenverlage Hans Böcker Stiftung. 40 p.

HOSSIEP, R. – FRIEG, P. (2008): Der Einsatz von Mitarbeiterbefragungen in Deutschland, Österreich und der Schweiz. Planung& Analyse, 6/2008, 55–59

ROSSI, P.H. – LIPSEY, M.W. – FREEMANN, H.E. (2004) Evaluation – A systematic approach (7th edition). Sage Publications Inc., Thousand Oaks. 469 pp.

STURGIS, P. – ROBERTS, C. – SMITH, P. (2014). Middle alternatives revisited: How the neither/nor response acts as a way of saying "I don't know"? Sociological Methods& Research, 43(1), 15–38.