Performance Journey Mapping: Evaluation of a Service Performance Assessment Framework

Angelika Höber
Department of Information Technologies & Business Informatics
CAMPUS 02 University of Applied Sciences (Austria)
email: angelika.hoeber@campus02.at (corresponding author)

Elisabeth Pergler
Department of Information Technologies & Business Informatics
CAMPUS 02 University of Applied Sciences (Austria)
email: elisabeth.pergler@campus02.at

Doris Weitlaner
Department of Information Technologies & Business Informatics
CAMPUS 02 University of Applied Sciences (Austria)
email: doris.weitlaner@campus02.at

Hans-Peter Grahsl
Department of Information Technologies & Business Informatics
CAMPUS 02 University of Applied Sciences (Austria)
email: hans-peter.grahsl@campus02.at

Abstract

This paper presents the evaluation results of a novel performance measurement framework, which has been designed for the particular needs of small and medium sized enterprises (SMEs). Prior research indicated that performance measurement is not much engaged in SMEs. The proposed framework aims to address this by creating bottom-up measurement systems. It connects the specific characteristics of SMEs with components from service management research and provides a set of tailored tools. Additional value with regard to system acceptance is obtained through contributions from goal-setting theory.

The framework as well as its tools have been evaluated in a first iteration of prototype implementation according to design science principles yielding promising results. Methodological triangulation was applied in order to ensure validity and reliability of the results from a pilot case study. In the course of the evaluation the framework was assessed from a quality, user and bird's-eye perspective.

Keywords
evaluation; performance measurement; service performance assessment; SMEs

1. Introduction

Contemporary performance measurement systems (PMS) are strongly influenced by the goods-dominant logic they have primarily been built for. Service industry hence lacks PMS which pay tribute to the particular characteristics of services. A set of alternate approaches has been proposed in the past, of which ITIL or SERVQUAL can be named as examples. In small and medium sized enterprises (SMEs) though only a poor application of performance measurement can be noted (Nudurupati et al., 2011). Reasons for this include: First, the complexity of these systems. High-level PMS require the implementation of highly sophisticated tools and/or techniques which are mostly not available in SMEs, not least because of their limited resources in regard to both time and costs. Second, the lack of acceptance-triggering characteristics such as ease-of use or usefulness. SMEs usually tend to prioritize activities connected to daily work and have no extra time for additional activities. More than this, PMS are often perceived as "a cause of bureaucratization and an obstacle to the flexibility of SMEs" (Garengo et al., 2005, p.30). Third, the small number of individuals. When decisions are depending on the perceptions and opinions of small groups it becomes crucial to have each employee's commitment. PMS however usually derive their performance goals from the company strategy (Bourne et al., 2000) without taking the staff's needs into account.

These problems have been addressed in the development of Performance Journey Mapping, a service performance assessment framework that connects the specific characteristics of SMEs with components from service management research. It aimed at providing a framework that is (i) easy to implement, (ii) possessing acceptance-triggering characteristics and (iii) fostering support by the individuals in the company. To this end, the Technology Acceptance Model (TAM) (Davis, 1989) was engaged for triggering acceptance. The Goal setting theory (GST) (Locke & Latham, 2002) should contribute to the generation of commitment. The proposed framework was finally designed as a bottom-up-approach which can easily be implemented.

This paper presents the results of the evaluation that focused on the question if the application of TAM and GST in the context of PMS development contributes to achieving the targeted outcomes. The evaluation was conducted according to the design science principles provided by Hevner et al. (2004). The framework was assessed from three different perspectives: the quality-, user- and bird's eye point of view. Methodological triangulation was used to ensure validity and reliability of the results. The findings are grounded on participant observation, questionnaires and formative evaluation.

The paper proceeds as follows: The next section outlines the scientific process. In the subsequent section the framework as well as its tools are outlined. Section 4 gives an overview of PMS evaluation concepts and discusses theoretical foundations. Section 5 details the evaluation concept and instruments. In section 6 findings are presented and discussed. The paper concludes with an outlook on the next steps.

2. Methodology

In order to ensure both, a scientific process and outcome, design science principles (Hevner et al., 2004) were applied to the development and evaluation of the performance assessment framework and the corresponding toolset. These principles have been transferred to the context of human-computer interaction by Hevner and Zhang (2011). The transfer resulted in an iterative approach consisting of three cycles:

- The relevance cycle, that focusses on practical usefulness of the artifact(s).
- The rigor cycle, that ensures contribution(s) to the knowledge base in the particular field of research.
- The internal design cycle, that connects methodological rigor and theoretical contributions to evaluation and design methods.

The relevance cycle in the present research project provides input regarding particular requirements of SMEs. Thus, the main evaluation criterion for the performance assessment framework is its ability to support and improve performance measurement activities in SMEs. It is fulfilled if the framework is accepted and applied by the individuals in a company, and results in a performance measurement system, that motivates individuals to fulfil their specific goals on the one hand, and on the other hand supports their performance measurement needs. In the course of the first artifact design iteration we conducted semi-structured interviews with SME representatives from different industries to obtain insights regarding their particular performance measurement needs as well as their currently implemented performance measurement methods and techniques. The interviews provide anecdotal evidence that (i) there is little consensus about what aspects and factors constitute the performance of service, and (ii) there is a lack of appropriate systematic approaches to conduct service performance analysis. In the context of the paper at hand, the two most relevant findings are the following:

- Few SMEs follow standardized (i.e. formalized and/or systematic) procedures to develop, operate and measure the performance of existing or novel services.
- Even SMEs that have defined methodologies and techniques do not employ them in everyday business due to SME-inherent reasons (time-to-market constraints, resource and/or capacity shortages etc.).

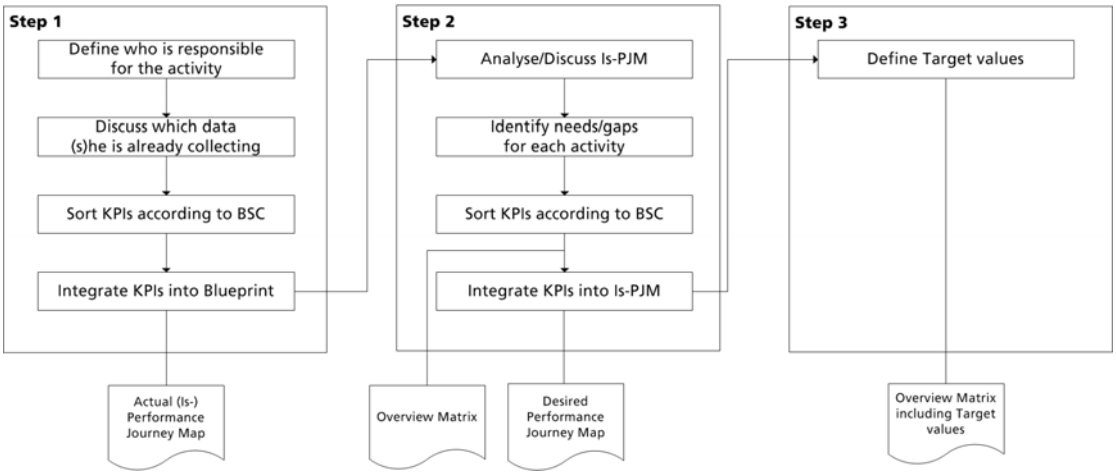
The rigor cycle is fed by two theoretical contributions from different fields of research. Firstly, the Technology Acceptance Model (TAM) (Davis, 1989) is used to improve and later evaluate acceptance and usage of the developed artifact. Second, Goal setting theory (GST) (Locke & Latham, 2002) is utilized to improve and later evaluate the motivational characteristics of the artifacts on an individual level. Moreover renowned mapping methods from service design were deployed to develop the visual tool supporting the performance assessment framework.

The internal design cycle consists of rapid artifact design and feedback iterations. Building and evaluation methods come from the rigor cycle, whereas immediate prototype tests are used to quickly obtain preliminary evaluation results from practical implementations. Early involvement of potential users of the framework enables continuous improvement of the artifact. A co-creation approach is applied to allow for user involvement at early stages. The velocity of iterations limits the applicability of quantitative evaluation methods. Method triangulation is applied to ensure validity and reliability of the iteratively obtained evaluation results based on small samples.

3. Performance Journey Mapping

Performance Journey Mapping is a bottom-up approach for designing PMS in SMEs. It was developed against the background of the specific characteristics of both SMEs and services. The procedure model is outlined in (Figure 1).

Figure 1. Outline of the three-step performance journey mapping process



The process starts with the collection of all performance measures already applied in an enterprise. Data are gathered based on a visual representation of the service delivery process, i.e. a Service Blueprint. Activities and processes on the firm side are in the center stage of the further course of action and are individually examined in a first iteration. The investigation is guided by two questions: "*Which stakeholders are affected? Which measures are already ascertained by them?*" The results are captured within a table and assigned to the appropriate dimensions of the Balanced Scorecard (BSC) – *finance*, *process*, *market*, and *potential* (Kaplan & Norton, 1992). The consequent table's content is merged with the Service Blueprint in the so-called *Performance Journey Map* (PJM). It shows (i) the coverage of the service process through performance measures and (ii) the (im-)balance within the BSC dimensions. Measurement gaps, i.e. single activities exhibiting no performance measures, or a lack of balance according to BSC, can easily be identified in the PJM.

Collecting requirements for future performance measurement constitutes the first task in Step 2. It is followed by a second iteration along the service delivery process addressing the question "*What (else) could or should be measured at that point?*" at each activity. As part of the discussion the *Performance Index* (PI) is utilized as pool of measures. Special attention is directed to the measurement gaps identified in the first iteration. New measures are, again, listed in a synoptic table, assigned to the BSC dimensions, and finally integrated into the PJM. As new measures have been added, methods for their collection and assessment need to be defined. Finally respective target values need to be defined. The final PJM provides an overview of the new PMS whereas the table serves as the basis for its operationalization.

The process described should be repeated regularly in order to (i) further develop the PMS and (ii) ensure the PMS' appropriateness for both, the current needs of SMEs and effective benchmarking with competitors in the long run. The adaptive nature of the PMS might induce a lack of historical (internal) comparability. However, this appears to be negligible in the light of the advantages derived from comparative performance benchmarking (Eccles, 1991).

3.1. Performance Journey Map

The PJM is a supporting tool of the framework. It is a visualization of the entire performance measures of a specific service along the service delivery process. It incorporates three perspectives which are of special interest for service performance measurement: (i) the service delivery process itself, (ii) the dimensions of the BSC, and (iii) the stakeholders within a company. The process representation in the form of a Service Blueprint gives an overview of the performance measurement and hence the distribution of the performance measures along the service process. The BSC dimensions provide additional insight into the measures'

allocation. By these means, dominant dimensions can be revealed and the creation of a balanced measurement simultaneously is encouraged.

In addition, the inclusion of stakeholders increases the prospect of the PMS' acceptance as employees decide for themselves whether they accept a measure and its corresponding benchmarking target or not (Wallace & Sheetz, 2014).

3.2. Performance Index

In the context of conceptualizing and designing the PJM a plethora of KPIs was identified based on a literature review including sources like Gladen (2011) and McNaughton et al. (2010). The resulting pool turned out to be too comprehensive in order to be helpfully applied during the joint development of a PJM. On that account, the collection of KPIs had to be substantially reduced. For that purpose the pool was filtered based on expert knowledge concerning the measures' aptitude in the SME context. The following aspects were crucial for the assessment: (i) meaningfulness in the service area, (ii) effort for collection/measurement, (iii) need for/availability of an ERP system, and (iv) the measure's mathematical/statistical complexity regarding its interpretability without expert know-how. This led to a PI consisting of a manageable number of KPIs that can be addressed while implementing a PJM in SMEs.

The additional classification of the remaining KPIs according to the BSC dimensions supports the pool's effective application. Moreover, the KPIs were binary annotated by reference to practically relevant attributes (e.g. the KPI's orientation, characteristics, ascertainment, or interpretability) in order to better support the purposeful selection of performance measures. The categorized and annotated PI is not assumed to be static. In contrast, it should be successively adapted and expanded based on the insights gained by its practical application.

4. Theory driven development

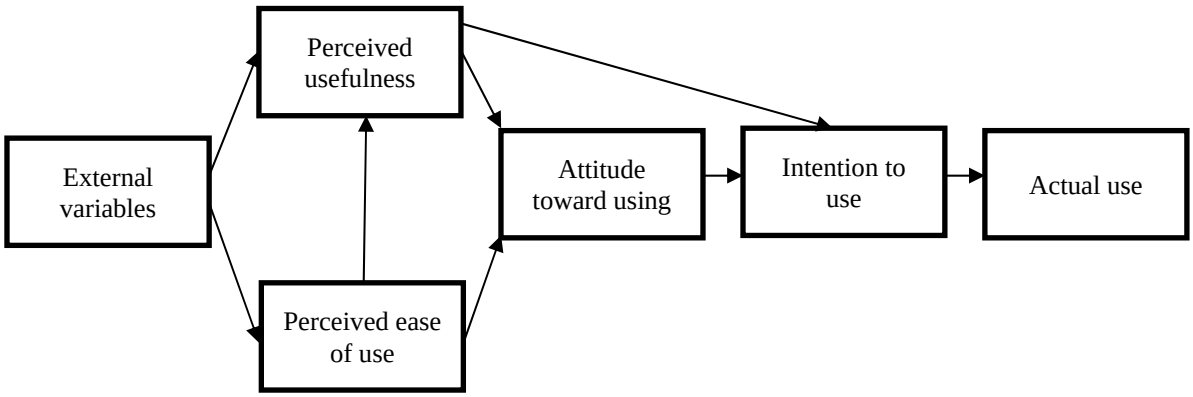
Literature provides a small number of proposals for the evaluation of PMS. Based on a literature review Jonsson and Lesshammer (1999) propose six crucial requirements (strategy, flow orientation, internal & external efficiency, improvement drivers, and simply & dynamic) for an overall manufacturing performance (OMP) system. Four of them are critical dimensions referring to what to measure while the other two represent characteristics which reflect on how to measure. By means of three case studies comprising medium- to large-sized manufacturers from different industries, the authors assessed the overall equipment efficiency measure of the PMS against their ideal requirements. A major finding was the common weakness of the employed PMS to hardly measure flow orientation or external effectiveness. Tangen (2004) developed a set of tools for the evaluation and revision of existing PMS. It contains a generally applicable classification scheme ranging PMS in five classes from "fully integrate" to "condemned". His key contribution is the so-called performance measurement progression map (PMPM) representing a guide which provides a systematic way of how to evaluate and revise an existing PMS in nine steps using selected methods and tools. An important aspect is that the PMPM is totally independent of the origin of an existing PMS and is solely focused to enhance the fulfilment of important PMS requirements in order to improve its overall performance.

Given the individual performance measurement needs of SMEs it appears reasonable to primarily evaluate the core of PMS design: the single measures. Several authors, e.g. Globerson (1985), emphasize one central requirement: performance measures must be derived from the company's strategy. This is hardly applicable for SMEs who mostly do not formulate a strategy (Garengo et al., 2005). Neely et al. (2000) provide a set of key characteristics for performance measures. They consolidate contributions from renowned

authors resulting in a list that can easily be adapted and applied to evaluate any set of performance measures.

Assessing the formal quality of PMS is not sufficient in SME environments. Various reasons for the poor take-up of PMS in SMEs have been discussed. To achieve successful use of PMS additional demands need to be considered. Acceptance and usage of the PMS are required for its effectiveness. Besides acceptance-triggering characteristics of the PMS itself it is necessary for the implementation of the framework that it is accepted as well. There exist numerous explanatory models of technology acceptance and usage, e.g. the Unified Theory of Acceptance and Use of Technology (Venkatesh et al., 2003) the model of information system success (DeLone & McLean, 1992) or TAM (Davis, 1989). The latter is renowned for its high explanatory power in different application scenarios. Wallace and Sheetz (2014) successfully validated a model derived from TAM in the context of software measures. The original TAM is focused on two factors that explain the behavioral intention to use a certain technology – perceived usefulness and perceived ease of use. (Figure 2) below outlines the suggested relationships among the constructs in TAM.

Figure 2. Technology Acceptance Model (TAM)



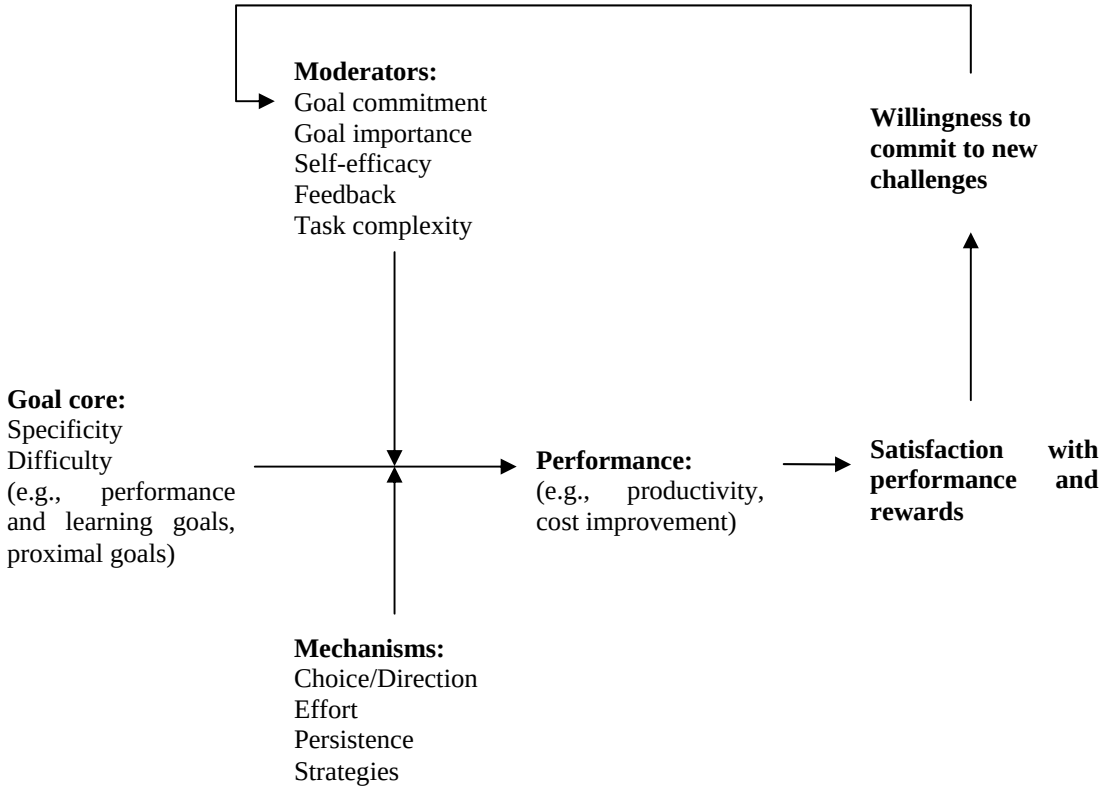
Source: Davis (1989)

The two basic TAM constructs, perceived usefulness and perceived ease of use were utilized as guidelines for the design of the performance assessment framework and its corresponding tools. The PJM itself as the visual tool within the framework serves as a parsimonious design of the entire PMS. The reduction to only one representation that is used for PMS creation and actual implementation will make it easy to learn interaction with the PMS. Moreover the simple design will enable employees to become experts in using the map quickly. The iterations that are necessary to dynamically adapt the PMS are simplified by the fact, that SMEs can do it themselves after an initial workshop where they learn the utilization of PJM.

Motivational aspects are required to enable performance improvements caused by the PMS. Among the plethora of motivational theories and models only a few focus on measures that can be set in organizations to increase motivation of employees to contribute to company objectives. For instance, Reiss' (2004) model of motivation, which is used for motive profiling focusses on people's basic desires, that influence everything they do in their lives. These desires are useful to understand and design incentive schemes but the model lacks causality. GST (Locke & Latham, 2002) provides a comprehensive set of goal characteristics and actual measures that will improve employee performance as depicted in (Figure 3). According to GST goals need to be specific and moderately difficult (neither too easy nor too

hard to achieve) to have a positive impact on performance. The effect is increased by commitment to goals, employee's perception of goal importance, self-efficacy, feedback on performance as well as task complexity. Mechanisms affecting the relationship between goal characteristics and performance are choice/direction, effort, persistence and strategies. Performance improvements further lead to satisfaction with performance and rewards that come along with it. This satisfaction is supposed to result in willingness to commit to new challenges that will positively affect the effects of motivational goals on performance in future. Goal-setting in general was found to have a positive impact on profitability in various industries (Terpstra & Rozell, 1994). Evidence for the effectiveness of different goal-setting measures was summarized in (Gardner, 2011). A recent simulation study on goal-setting in deadline-oriented processes suggests that motivational goals result in improved operational performance (Doerr & Gue, 2013).

Figure 3. Elements of goal-setting theory and the high-performance cycle (Locke & Latham, 2002)

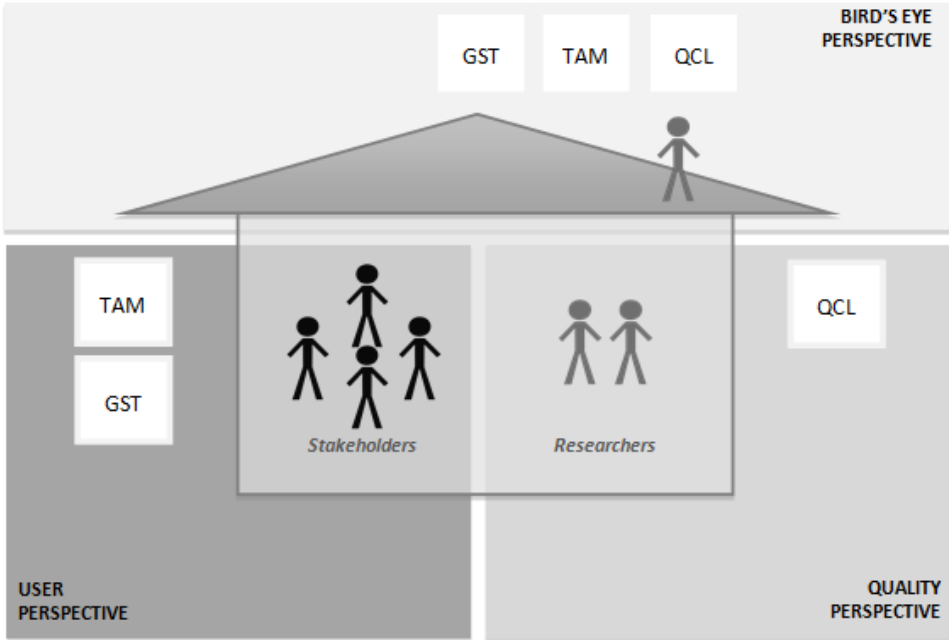


GST was implemented in the performance measurement framework in terms of stakeholder involvement. Employees who perform certain activities in the service delivery process are engaged to participate in the PMS creation. Their particular needs and requirements are integrated in the actual PJM. Benchmarking goals are set in the course of the workshop and include a holistic view on the entire service delivery process. Employees can, thus, influence and co-determine levels of specificity and difficulty of goals that affect them. The holistic perspective is suggested to foster perceptions of both, goal importance and goal commitment. The presence of the PJM in the workplace enables continuous observations of performance improvements as well as immediate feedback on goal-achievements, e.g. increased productivity or improved customer satisfaction.

5. Evaluation instrument development

Performance Journey Mapping was evaluated in a pilot project where the workshop concept was complemented by an evaluation concept. The single evaluation measures were executed at different points in time, in and after the workshop. The evaluation aimed at answering whether the application of TAM and GST did contribute to the development of a PMS that meets the defined goals and at the same time fulfils renowned PMS quality criteria as proposed by Neely et al. (2000). The results from the evaluation should feed back to the further development of PJM. The three addressed evaluation dimensions have been introduced in the previous chapters: (i) TAM, (ii) GST and (iii) a Quality criteria list (QCL) based on the listing provided by Neely et al (2000). To ensure valid and reliable results the dimensions have been assessed from different angles: quality, user and bird's eye perspective (Figure 4). The *quality perspective* corresponds to the researcher's view. It shows results from the perspective of the participating scientists and evaluates in a formal and formative way. The *user perspective* represents the users of the PMS. They are the addressees of the system and the actual experts (Kromrey, 2009). In the given evaluation the users are the single stakeholders of the service delivery process within a company. The framework is applied in a highly interactive workshop and its outcome depends on collaborative processes, both between the researchers and the users and amongst the stakeholders themselves. Thus it does not suffice to evaluate the perspectives in an isolated way. The *bird's eye perspective* is used to retrieve data in a collaborative setting and at the same time to identify potential problems or risks in the interactive setting itself.

Figure 4. Evaluation dimensions and perspectives



5.1. Evaluation dimension 1: Quality criteria list (QCL)

Starting from Neely's (Neely et al. 2000) listing of desirable characteristics some adaptations were made to obtain a useful formal evaluation basis. Two items which could not be evaluated in a design phase were removed, another three items (B1-B3) from a more recent contribution by the same author (Neely et al., 2005) were added and one custom item (C1) with special interest for the PJM was appended. The resulting list can be found in (Table 1).

Table 1. Quality criteria list

	Criterion	Subject
A1	Performance measures should enable/facilitate benchmarking.	Comparability
A2	Performance criteria should be directly under the control of the evaluated organizational unit.	Responsibility
A3	Objective performance criteria are preferable to subjective ones.	Objectiveness
A4	Non-financial measures should be adopted.	Non-Finance
A5	Performance measures should be simple and easy to use.	Comprehensibility
A7	Performance measures should stimulate continuous improvement rather than just monitor.	Improvement
B1	Have all the appropriate elements (internal, external, financial, non-financial) been covered?	Balance along BSC dimensions
B2	Have the measures been integrated, both vertically and horizontally?	Balance along company levels
B3	Do any of the measures conflict with one another?	Absence of conflict
C1	Performance measures are distributed along the service blueprint	Balance along service blueprint

The QCL was evaluated from quality- and bird's eye perspective. In the quality perspective an internal formative evaluation was performed subsequent to the workshop: two checklists were used to assess the single performance measures of the PJM. (The wording for the criteria had been changed in such way that it enabled a judgement of the single measures). Checklist 1 covered the criteria assessing the uniform distribution of the measures (B1, B2, C1): the single measures were assigned to one (or more) BSC dimension(s), one (or more) company levels and one blueprint activity. By calculating the sums for the single dimensions/ company levels/activities an overview of the balance was gained. Checklist 2 was a matrix with criteria in columns and measures in lines; it required a yes/no-judgement for each measure/criteria-combination.

In the bird's eye perspective participant observation was applied. For this purpose the criteria were operationalized as subjects. Throughout the workshop two researchers were observing whenever a subject was addressed. Their observations were documented in the form of field notes. In addition the users were asked for their explicit judgement in the collaborative setting: at the end of the workshop every participant was given a set of cards, each containing one criterion (rephrased from question to statement to enable judgement). In the light of the measures defined on the final PJM all participants were asked to individually sort their set of cards and to return only those they went along with.

5.2. Evaluation dimension 2: Technology Acceptance Model (TAM)

Davis (1989) provides a questionnaire consisting of eleven questions to assess perceived usefulness and perceived ease of use. One of the questions is tightly connected to IT and cannot be answered meaningfully in a PMS context. It was thus not used for the evaluation. The utilized list is contained in (Table 2).

Table 2. TAM questionnaire

	Proposition	Construct
T1	Using x in my job would enable me to accomplish tasks more quickly.	Usefulness
T2	Using x would improve my job performance.	
T2	Using x would improve my productivity.	
T3	Using x would enhance my effectiveness on the job.	
T4	Using x would make it easier to do my job.	
T5	I would find x useful in my job.	
T6	Learning to operate x would be easy for me.	Ease of use
T8	My interaction with x would be clear and understandable.	
T9	I would find x to be flexible to interact with.	
T10	It would be easy for me to become skillful at using x.	
T11	I would find x easy to use.	

The TAM was assessed from user- and bird's eye perspective. In the user perspective a simple questionnaire was engaged containing the ten propositions rateable by means of a five-point-scale ranging from "likely" to "unlikely". After the workshop the participants (on company side) were asked to answer the questionnaire. In the bird's eye perspective the same participants were asked to rate the PJM by posting sticking dots on a flipchart with a scale between a happy and a sad smiley. The flipchart contained two scales, one entitled "Usefulness" and the other "Ease of use". The activity was executed twice, first to rate the Is-PJM right after its creation (in the middle of the workshop) and a second time to rate the final PJM (at its end). Participant observation was used throughout the workshop to monitor any addressing of one of the basic constructs.

5.3. Evaluation dimension 3: Goal setting theory (GST)

Locke and Latham (2002) discuss five "moderators": factors which increase the positive impact of properly defined goals. One of them (task complexity) can hardly be evaluated seriously from an external view, it was thus not used in the evaluation. The remaining four moderators were operationalized as questions. In order to gain deeper insights two moderators have been transferred into two respectively three questions. (Table 3) gives an overview.

Table 3.Goal setting theory questions

	Question	Moderator
G1	I am ready to show large commitment in order to achieve the goals that have been defined for me.	Goal commitment
G2	I believe that the goals that have been defined for me are important.	Importance
G3	I believe that the goals that have been defined for me do contribute to the company's success.	
G4	I believe that I am able to achieve the goals that have been defined for me.	Self-efficacy
G5	I believe that with the help of the new PMS I will be informed if I am on the wrong track.	Feedback
G6	I believe that with the help of the new PMS I will get feedback on whether I am doing a good or bad job.	
G7	I believe that with the help of the new PMS I will get feedback when I am doing a good job.	

The moderators were assessed from user- and bird's eye perspective. In the user perspective a simple questionnaire was deployed. It contained the seven questions with a four-point-assessment scale ranging from "applies fully" to "does not apply at all". The participants on company side were asked to fill in the questionnaire after the workshop. In the bird's eye perspective participant observation monitored whether and when the moderators were addressed during the workshop.

6. Results

In the evaluated pilot project a PMS was designed for a SME that offers IT-services. The executive manager and ten employees participated in the workshop, including two project managers, the marketing executive and software development staff on different professional levels. The research team consisted of two workshop moderators, one person in charge of the PI and two more executing the participant observation. The workshop took place in the company's meeting room on a workday morning and had a firm-driven timely limitation of two hours.

6.1. Quality criteria list (QCL)

For the formative evaluation Checklist 1 was filled based on the PJM. Balance in regard to (i) BSC, (ii) company level or (iii) service blueprint is not assumed when (i) one BSC dimension has non or twice as many measures than another one, (ii) one company level has 25% more measures than the other one, and (iii) more than 25% of the blueprint activities have no measure at all. Checklist 2 was first processed by three researchers independently. In the succeeding consolidation a positive judegment was only accepted when at least two researchers had confirmed it. The cards returned by the participants were sorted according to the printed criteria and counted. The results of Checklist 2 and the numbers of the returned cards were computed on a quartile basis: results in the upper quartile were assigned a *plus (+)*, results in the lower quartile a *minus (-)*, for results in the interquartile range a *null (0)* was allocated. By means of selective coding the field notes were analyzed, using the subjects as predefined codes. This led to an overview of the frequency of mention of the single subjects as well as their drivers. The results were confronted in an overview and sorted according to their frequency.

Five out of ten criteria exhibit congruent results in both perspectives with either *minus* or *plus* (as can be seen in Table 4): *Comprehensibility* and *Absence of conflict* gained a *plus*. *Balance along the service blueprint* and *Balance along the company levels* reached a *minus. Non-Finance* was judged with *null*. The remaining five criteria reveal a slight discrepancy (one being *plus* or *minus*, the other *null*): *Responsibility*, *Improvement* and *Objectiveness* obtain a *null* in the quality and a *plus* in the bird's eye perspective. *Comparability* gained a *null* in the quality and a *minus* in the bird's eye perspective. The *Balance along BSC dimensions* concludes with a *minus* in the quality and a *null* in the bird's eye perspective.

6.2. Technology Acceptance Model (TAM)

The flipchart scales containing the sticking dots were photographed and subdivided into four quarters to determine the position of the dots. In the next step the two images were juxtaposed to determine possible shifts and trends. A positive trend was documented with a *plus*, a negative trend with a *minus*. To evaluate the questionnaire first the arithmetic means were classified in a quartile structure assigning *plus, minus* or *null* values. Then a final result was gained by calculating the arithmetic means of the affiliated questions per factor and valuing them in the same way. The selective coding of the field notes included the two TAM factors as predefined codes.

In the user perspective four out of eleven questionnaire requests resulted in a *plus* (T1, T4, T5, T11), the remainder reached a marginal *null*, missing the upper quartile by less than 0.6 points. The final result for both *Usefulness* and *Ease of use* resulted in *plus* values. The bird's eye perspective featured the same results: both factors displayed a shift with a positive trend and were thus rated with a *plus*.

6.3. Goal setting theory (GST)

The questionnaire was evaluated by classifying the arithmetic means, again in a quartile structure resulting in *plus*, *minus* or *null*. For the moderators with more than one question the arithmetic means were calculated and judged in the same way. The four moderators were also integrated as codes in the selective coding of the field notes.

In the user perspective one question (G5) was rated with a *null*, the remaining six questions scored a *plus.* The final results for all four moderators resulted in *plus.* The field notes revealed a broad range in the frequency of the moderators with *Importance* in a clear top position, followed by *Feedback*, *Goal Commitment,* and *Self-efficacy*.

Table 4.Overview of results

	Quality	User	Bird's eye	
			Participants' evaluation	Frequency
Quality criteria list (QCL)				
Responsibility	**0**		+	22
Improvement	**0**		+	13
Non-Finance	**0**		**0**	8
Balance along BSC dimensions	**-**		**0**	8
Comparability	**0**		-	7
Objectiveness	**0**		+	5
Balance along service blueprint	-		-	5
Comprehensibility	+		+	3
Absence of conflict	+		+	0
Balance along company levels	-		-	0
Technology Acceptance Model (TAM)				
Usefulness		+	+	3
Ease of use		+	+	2
Goal setting theory (GST)				
Importance		+		13
Feedback		+		6
Goal commitment		+		3
Self-efficacy		+		1

The evaluation yielded predominantly positive ratings: out of 28 individual evaluations15 resulted in a *plus*, 6 in a *null* and 7 in a *minus*. A large share of the findings shows congruence across the different perspectives. For the TAM factors and all GST moderators exclusively positive ratings were yielded. The evaluation along the QCL shows more diverse results but still with a good amount of congruence. *Comprehensibility* and *Absence of conflict* received positive results. Conflicts between single measures are more likely to arise when measures are defined independently for different departments. In the given case the company was not segmented into departments and the measures were defined in presence and with participation of nearly all stakeholders. The latter may also be considered in regard to *Comprehensibility* as the measures' wordings were defined in that setting. Negative ratings were yielded for the criteria connected to balance. *Balance along company levels* received negative ratings from both users and researchers. Each single measure has been assigned to a stakeholder on operational level, in part also to stakeholders on management level what resulted in an overhang on operational side. Considering the nature of SMEs with a rather small share engaged in management the definition used for *Balance along company levels* may need to be revised in the next iterations. *Balance along service blueprint* was also rated negative from both users and researchers. The created PJM reveals an emphasis on the left half of the service blueprint (16 out of 23 measures). The timely limitation of the workshop could have promoted this situation as the identification of measures had been started on the left side and the final balance was not discussed, due to lack of time. Furthermore single focal points could be noted on activities connected to customer contract meetings. This is in accordance with *Balance along BSC dimensions* which was rated with *minus* respectively *null*. The dimensions Market (9) and Finance (8) dominate over Potential (3) and Process (4), which affects also

Non-Finance that resulted in a *null*. *Responsibility*, *Improvement* and *Objectiveness* gained better results in the bird's eye perspective than in the formal evaluation. *Comparability* in turn received a *null* in the formal evaluation and a completely negative evaluation in the collaborative setting. The frequency of mention shows that *Responsibility* was a major issue in the workshop. Interestingly enough the field notes displayed an overlap of *Responsibility* (QCL) and *Importance* (GST), which frequently appeared at the same time. Reasons might be found in the design of Performance Journey Mapping: in the pilot project the single measures were proposed and defined by their stakeholders. They showed a confidence in the importance of those measures and were at the same time responsible for them.

7. Conclusion & Outlook

The evaluation results show that the application of TAM and GST did contribute to develop a PMS that is characterized by Usefulness and Ease of Use. Deficiencies could be noted in regard to renowned quality criteria for PMS, especially in respect of balance. First lessons learned can be outlined:

- The dimensions of the BSC as well as the concept of the Service Blueprint are essential basics for the creation of a PJM. Participants repeatedly pointed out that they did not understand the meaning of the BSC dimensions. Besides there was no common understanding of what a "Performance measure" is or can be.
- The timely restriction for the workshop was too tight. While the progress in the first half of the workshop was promising and participants showed growing awareness, the second half was characterized by increasing unrest.
- There are several indications that a top-down dominance was present in the pilot project. The vast majority of documented interactions were conducted by the executive management or the project management (53 out of 65). This is also reflected in the distribution of the measures, their focal points and dominant BSC dimensions.

Performance Journey Mapping simultaneously fosters and requires support from all stakeholders of a service. In the next iterations it needs to be investigated if the prevalence of management could be connected to the lacking basics and the difficulties in verbalizing measures on side of the employees. Furthermore it has to be examined whether a larger time frame can contribute to a more balanced outcome. The preliminary definitions of balance need to be critically reflected and refined.

References

Bititci U.S., Carrie A.S., McDevitt L. (1997). "Integrated performance measurement systems: A development guide". *International Journal of Operations & Production Management*, 17(5), 522-534.

Bourne M., Mills J., Wilcox M., Neely A., Platts K. (2000). "Designing, implementing and updating performance measurement systems". *International Journal of Operations & Production Management*, 20(5), 754-771.

Davis F.D. (1989). "Perceived usefulness, perceived ease of use, and user acceptance of information technology". *MIS Quarterly*, 13(3), 319-340.

DeLone W., McLean E. (1992). "Information system success: The quest for the dependent variable". *Information Systems Research*, 3(1), 60-95.

Doerr K.H., Gue K.R. (2013). "A performance metric and goal-setting procedure for deadline-oriented processes". *Production & Operations Management*, 22(3), 726-738.

Eccles R.G. (1991). “The performance measurement manifesto”. *Harvard Business Review*, 69(1/2), 131-137.

Gardner A.C. (2011). “Goal setting and gainsharing: The evidence on effectiveness”. *Compensation & Benefits Review*, 43(4), 236-244.

Garengo P., Biazzo S., Bititci U.S. (2005). “Performance measurement systems in SMEs: A review for a research agenda”. *International Journal of Management Reviews*, 7(1), 25-47.

Gladen W. (2011). *Performance Measurement: Controlling mit Kennzahlen (In German, English title: Performance Measurement: Controlling with Key Figures)*. Wiesbaden: Gabler.

Globerson S. (1985). “Issues in developing a performance criteria system for an organization”. *International Journal of Production Research*, 23(4), 639-646.

Hevner A., March S.T., Park J., Ram S. (2004). “Design Science in Information Systems Research”. *MIS Quarterly*, 28(1), 75-105.

Hevner A., Zhang P. (2011). “Introduction to the AIS THCI special issue on design research in human-computer interaction”. *AIS Transactions on Human-Computer Interaction*, 3(2), 56-61.

Jonsson P., Lesshammar M. (1999). “Evaluation and improvement of manufacturing performance measurement systems – the role of OEE”. *International Journal of Operations & Production Management*, 19(1), 55-78.

Kaplan R.S., Norton D.P. (1992). “The balanced scorecard: Measures that drive performance”. *Harvard Business Review*, 70(1), 71-79.

Kromrey H. (2009). *Empirische Sozialforschung: Modelle und Methoden der standardisierten Datenerhebung und Datenauswertung (In German, English title: Empirical Social Research: Models and Methods of Standardized Data Collection and Evaluation)*. Stuttgart: Lucius & Lucius.

Locke E.A., Latham G.P. (2002). “Building a practically useful theory of goal setting and task motivation: A 35-year odyssey”. *American Psychologist*, 57(9), 705-717.

McNaughton B., Ray P., Lewis L. (2010). “Designing an evaluation framework for IT service management”. *Information & Management*, 47(4), 219-225.

Neely A., Gregory M., Platts K. (2005). “Performance measurement system design: A literature review and research agenda”. *International Journal of Operations & Production Management*, 25(12), 1228-1263.

Neely A., Mills J., Platts K., Richards H., Gregory M., Bourne M., Kennerley M. (2000). “Performance measurement system design: Developing and testing a process-based approach”. *International Journal of Operations & Production Management*, 20(10), 1119-1145.

Nudurapati S.S., Bititci U.S., Kumar V., Chan F.T.S. (2011). “State of the art literature review on performance measurement”. *Computers & Industrial Engineering*, 60(2), 279-290.

Reiss S. (2004). “Multifaceted nature of intrinsic motivation: The theory of 16 basic desires”. *Review of General Psychology*, 8(3), 179-193.

Tangen S. (2002). “Understanding the concept of productivity”. *Proceedings of the Asia Pacific Industrial Engineering and Management Systems Conference (APIEMS2002)*, December 18-20, Taipei, Taiwan.

Tangen S. (2004). *Evaluation and Revision of Performance Measurement Systems*. Doctorial Thesis, The Royal Institute of Technology, Stockholm, Sweden.

Tangen S., Johansson B. (2000). *Produktivitet – undersökning på 6 företag (In Swedish, English title: Productivity – A study of 6 Companies)*. Internal Report, The Royal Institute of Technology, Stockholm, Sweden.

Terpstra D.E., Rozell E.J. (1994). “The relationship of goal setting to organizational profitability”. *Group & Organization Management*, 19(3), 285-294.

Venkatesh V., Morris M.G., Davis G.B., Davis F.D. (2003). “User acceptance of information technology: Toward a unified view”. *MIS Quarterly*, 27(3), 425-478.
Wallace L.G., Sheetz S.D. (2014). “The adoption of software measures: A technology acceptance model (TAM) perspective”. *Information & Management*, 51(2), 249-259.